(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
30.07.1997 Bulletin 1997/31

(51) Int Cl.⁶: **C12N 15/00**

(21) Application number: 97100117.7

(22) Date of filing: 07.01.1997

---

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV RO SI

(30) Priority: 05.01.1996 US 9861

(71) Applicant: HUMAN GENOME SCIENCES, INC.
Rockville, MD 20850-3338 (US)

(72) Inventors:
• Kunsch, Charles A.
Gaithersburg, Maryland 20882 (US)

• Choi, Gil H.
Rockville, Maryland 20850 (US)
• Barash, Steven C.
Rockville, Maryland 20850 (US)
• Dillon, Patrick J.
Gaithersburg, Maryland 20879 (US)
• Fannon, Michael R.
Silver Spring, Maryland 20906 (US)
• Rosen, Craig A.
Laytonsville, Maryland 20882 (US)

(74) Representative: VOSSIUS & PARTNER
Postfach 86 07 67
81634 München (DE)

---

(54) **Staphylococcus aureus polynucleotides and sequences**

(57) The present invention provides polynucleotide sequences of the genome of Staphylococcus aureus, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynu- cleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

EP 0 786 519 A2

Description

Th present invention relates to the field f molecular biology. In particular, it relates to, among other things, nucleotid sequences of *Staphylococcus aureus*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequenc s, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., The *Staphylococcus* as a Molecular Genetic System, Chapter 1, pgs. 1-37 in MOLECULAR BIOLOGY OF THE *STAPHYLOCOCCI*, R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus*, a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

Human Health and S. Aureus

Staphylococcus aureus is a ubiquitous pathogen. (See, for instance, Mims *et al.*, MEDICAL MICROBIOLOGY, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, eyelid infections, food poisoning, joint infections, neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

Burns

Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

Cellulitis

Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyrogenes*. Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic *streptococci*. It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

Eyelid infections

S. aureus is the cause of styes and of sticky eye" in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria. Although the toxins are known, their mechanism of action is not understood.

Joint infections

S. aureus infects bone joints causing diseases such osteomyelitis.

Osteomyelitis

S. aureus is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in

children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphysial growth plates in the nd of long, growing bones.

Skin infections

S. *aureus* is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections. Recurrent infections of the nasal passages plague nasal carriers of S. *aureus*.

Surgical Wound Infections

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient. S. aureus is the most important causative agent of infections in surgical wounds. S. *aureus* is unusually adept at invading surgical wounds; sutured wounds can be infected by far fewer S. *aureus* cells then are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe S. *aureus* septicaemia. Invasion of the blood stream by S. *aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

Scalded Skin Syndrome

S. *aureus* is responsible for "scalded skin syndrome" (also called toxic epidermal necrosis, Ritter's disease and Lyell's disease). This diseases occurs in older children, typically in outbreaks caused by flowering of S. *aureus* strains produce exfoliation(also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the diseases. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

Toxic Shock Syndrome

Toxic shock syndrome is caused by strains of S. *aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by S. *aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxaemia and septicaemia, and can be fatal.

Nocosomial Infections

In the 1984 National Nocosomial Infection Surveillance Study ("NNIS") S. *aureus* was the most prevalent agent of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

Resistance to drugs of S. aureus strains

Prior to the introduction of penicillin the prognosis for patients seriously infected with S. *aureus* was unfavorable. Following the introduction of penicillin in the early 1940s even the worst S. *aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of S. *aureus* did not take long, however. Most strains of S. *aureus* encountered in hospital infections today do not respond to penicillin; although, fortunately, this is not the case for S. *aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of S. *aureus* produce a lactamase which converts penicillin to pencillinoic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episo-mally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in S. *aureus*. These compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in S. *aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of S. *aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistanc in *S. aureus* has been elucidated (See Lyon *et al.*, *Microbiology Reviews* 51: 88-134 (1987)). G nerally, resistance is m diated by plasmids, as noted abov regarding p nicillin resistance; how v r, several stable forms of drug resistance have been observed that apparently involve integration of a r sistance element into the *S. aureus* genome itself.

Thus far each new antibiotic gives rise to resistance strains, stains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

Molecular Genetics of *Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

Most genetic studies of *S. aureus* have been carried out using the the strain NCTC8325, which contains prophages psi11 psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci*, consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in. MOLECULAR BIOLOGY OF THE *STAPHYLOCOCCI*, R.P. Novick, Ed., VCH Publishers, New York, (1990) The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J. , *Antimicrobial Chemotherapy* 21 Suppl C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections. There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-5,191.

The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

The nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS:1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to:magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, pref rably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs," fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs," and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs."

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention can from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vacciniating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the Dfs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following:wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a)contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b)determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genom s will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

FIGURE 1 is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

FIGURE 2 is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al., Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences,* 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR") for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3..

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-5,191; or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-5,191" refers to any portion of the SEQ ID NOS:1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs"), expression modulating fragment (EMFs") and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

As discussed in detail below, the information provided in SEQ ID NOS:1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-5,191. However, once the present invention is made available (*i.e.,* once the information in SEQ ID NOS:1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

As discussed elsewhere hererin, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC°).

5   The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191, in a form which can be readily

10   used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate

15   an identity score of polynucleotides compared to one another.

COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide

20   sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, Òprovided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.,* a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical

25   to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (*e.g.,* a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer

30   readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD- ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture com-

35   prising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium

40   to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer

45   readable medium. The sequence information can be represented in a word processing text file, formatted in commercially- available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.,* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

50   Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

55   The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.,* J. Mol. Biol. 215:403410 (1990)) and BLAZE (Brutlag *et al.,* Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein ar   protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in th   production of commercially useful metabolites.

The pr sent invention furth r provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, comm rcially important frag-

5   ments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available

10   computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the

15   present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer- based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which

20   match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

25   As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and

30   protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited

35   to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the

40   target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can

45   readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104

50   are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes

55   appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for

accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-5,191, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generated the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of Staphylococcus aureus genomic DNA. Thus, given the availability of SEQ ID NOS:1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

In Tables 1 and 2, column six, lists the Reference" for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the numenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name" of the matching sequence; column eight provides the BLAST identity" score from the comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair" identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions

1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.,* possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity" of the highest scoring s gment pair" in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host in examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS:1-5,191, with a sequence from another isolate of the same species.

Furthermore, to accomodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.,* sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particular *Staphylococcus aureus*. Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Sta-*

phylococcus aureus, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control g ne expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee et al., Nucl. Acids Res. 6: 3073 (1979); Cooney et al., Science 241: 456 (1988); and Dervan et al., Science 251: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, J. Neurochem. 56: 560 (1991) and OLIGODEOXYNUCLE-OTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the Staphylococcus aureus genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the Staphylococcus aureus genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacl, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the Staphylococcus aureus genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. et al., BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the Staphylococcus aureus genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polpeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography.

Th polypeptides and proteins of the present invention also can be purified from cells which have be n altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to xpress a desir d polypeptide or prot in when th cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in th art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial"defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alphafactor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus, E. coli, B. subtilis, Salmonella typhimurium* and various species within the genera *Pseudomonas, Streptomyces,* and *Staphylococcus*. Others

12

may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available form Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, Cell 23: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes Staphylococcus aureus polypeptides which are useful as immuno-diagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outermembrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., Mol. Microbiol. 13, 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., J. Bacteriol. 174, 7345-7351; 1992)). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. Lipoproteins in bacteria. J Bioenerg. Biomembr. 22, 451-471; 1990).

It well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as S. pyogenes, S. mutans, E. faecalis, S. pneumoniae, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity. ASM News 62, 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

amin d. The amino acid sequence of this region is L-P-X-T-G-X, wher X is any amino acid.

Amino acid sequ nce similarities to proteins of kn wn function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins ar well known in the art and include "lipoprotein", "periplasmic", or "antigen".

An algorithm for s lecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORFs which are predicted to be outermembrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS:5,192 to 5,255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS:5,192-5,255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for xample the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS:5,192-5,255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding ORF listed in Tables 1,2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORFs identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of amino acid sequences shown in the sequence listing as SEQ ID NOS:5,191-5,255; or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Learner, R. A. (1983) "Antibodies that react with predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate S. aureus specific antibodies include: a polypeptide comprising peptides shown in Table 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated S. aureus proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides:

specificity of antigen-antibody interaction at the l vel of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further describ d in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson t al., supra;
5   Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al., supra. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the
10   epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Per-alkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such
15   peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a per-alkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outermembrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the
20   identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO:5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are
25   polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are sub-stantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more
30   substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity be-tween reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus*, of the
35   fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the
40   sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybrid-ization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which prossess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among
45   especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-5,191 or from
50   a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*, PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

55   When using primers derived from SEQ ID NOS:1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-5,191, or from a nucl otid sequence having an aforementioned id ntity to a sequ nce of SEQ ID NOS:1-5,191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (e.g., hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacterias which are closely related to Staphylococcus aureus.

ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the Staphylococcus aureus ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESES, Tramper et al., Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in Staphylococcus aureus can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-5,191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts et al., Symbiosis 21: 79 (1986) and Voragen et al. in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker et al., Eds., American Chemical Society Symposium Series 389: 93 (1989) .

The metabolism of sugars is an important aspect of the primary metabolism of Staphylococcus aureus. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger et al., Biotechnology 6(A), Rhine et al., Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir et al., Biotechnology Letters 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis et al., beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett et al., Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for

quantitativ determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al., Biochem. et Biophysica. Acta.* 872: 83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.,* Biotechnology, The Textbook of Industrial Microbiology, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, Starch 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.,* Acid Proteases Structure Function and Biology, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.,* Industrial Enzymes, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.,* Report Industrial Enzymes by 1990, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.,* Philosophical Transactions of the Chiral Society of London 310:227 (1985) and Poserke, Journal of the American Oil Chemist Society 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.,* Recent Advances in the Generation of Chiral Intermediates Using Enzymes, CRC Press, Boca Raton, Florida (1990). The following reactions catalyzed by enzymes are of interest to organic chemists:hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.,* Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus.* These include Sau3A and Sau96I.

2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al., J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256: 495-497 (1975)), the trioma technique, the human B- cell hybridoma technique (Kozbor *et al., Immunology Today*

<u>4</u>: 72 (1983), pgs. 77-96 of Col *et al.,* in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985)).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogen polypeptid . Methods for immunization ar w II known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al., Exp. Cell Res.* <u>175</u>: 109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above- described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.,* J. Histochem. Cytochem. 18:315 (1970); Bayer, E. A. *et al.,* Meth. Enzym. 62:308 (1979); Engval, E. *et al.,* Immunol. 109:129 (1972); Goding, J. W. J. Immunol. Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro, in vivo,* and in situ assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.,* "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.,* Meth. Enzym. 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro, in vivo,* and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs,antigens or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the Dfs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.,* Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry; PCT publication WO95/32291, and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based

on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be r adily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain th necessary r agents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises:(a) a first container comprising one of the Dfs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following:wash reagents, reagents capable of detecting presence of a bound DF, antigen or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed Dfs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of:

(a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the Staphylococcus aureus genome; and
(b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.,* Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.,* Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.,* Nucl. Acids Res. 6:3073 (1979); Cooney *et al.,* Science 241:456 (1988); and Dervan *et al.,* Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effectiv in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

5. Pharmaceutical C mpositions and Vaccin s

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic Staphylococcus aureus polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e. g., human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

ILLUSTRATIVE EXAMPLES

LIBRARIES AND SEQUENCING

5

1. Shotgun Sequ n ing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, Genomics 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P_0, that any given base in a sequence of size L, in nucleotides, is not sequenced after

10 a certain amount, n, in nucleotides, of random sequence has been determined can be calculated by the equation $P_0 = e^{-m}$, where m is L/n, the fold coverage." For instance, for a genome of 2.8 Mb, m=1 when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, $P_0 = e^{-1} = 0.37$. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivilent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage,

15 approximately 37% of a polynucleotide of size L, in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a .2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length, G, is determined by the equation $G = Le^{-m}$, and the average gap size, g, follows the

20 equation, g = L/n. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a poly-nucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, Genomics 2: 231 (1988).

2. Random Library Construction

25

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

Staphylococcus aureus DNA was prepared by phenol extraction. A mixture containing 600 ug DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson

30 Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500 ul TE buffer.

To create blunt-ends, a 100 ul aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 ul BAL31 buffer . The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100 ul TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting

35 temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20 ul of TE buffer for ligation to vector.

A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contained 2 ug of DNA fragments, 2 ug pUC18 DNA (Pharmacia) cut with Smal

40 and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20 ul TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, etc. The portion of the gel containing v+i DNA was excised and the v+i DNA was recovered and resuspended

45 into 20 ul TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37° C in a reaction mixture (50 ul) containing the v+i linears, 500 uM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linears were dissolved in 20 ul TE. The final ligation to produce circles was carried out in a 50 ul reaction containing 5 ul of v+i linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was

50 stored at -20°C.

This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, E.coli host cells deficient in all recombination and restriction functions (A. Greener, Strategies 3 (1):5 (1990)) were used to prevent rearrangements,

55 deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100 ul aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 ul aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1 ul aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation
5   mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl$_2$ (1 M), and 1 ml MgSO$_4$/100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10 ul aliquot of transformation.

10   All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

3. Random DNA Sequencing
15   ·

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime → 3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.
20   Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200 ul) containing 50 ug DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuges over a 10- 40% sucroce gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation . One ul of fragments was used with 1 ul of DASHII vector (Stratagene) in
25   the recommended ligation reaction. One ul of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract Phage were plated directly without amplification from the packaging mixture (after dilution with 500 ul of recommended SM buffer and chloroform treatment). Yield was about 2.5x10^9 pfu/ul.

An amplified library was prepared from the primary packaging mixture according to the manufactureer's protocol.
30   The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1x10^9 pfu/ml.

Mini-liquid lysates (0.1ul) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready
35   Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of AB 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed
40   using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

45   ### 4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing
50   primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e.., one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.
55   Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-prim rs and dye- terminators with approximately equal fidelity, although plasmid templates generally give longer usable s quences.

5   Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence

10   itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

INFORMATICS

15

1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al., Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System*

20   *Sciences,* IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow whereever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based

25   on a Unix platform, it was necessary to design and implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

2. Assembly

30   An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was enployed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 10^4 fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements.

35   Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched

40   end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repet-itive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of align-ments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information

45   coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain ranged of base pairs (definable for each clone based on the known clone size range for a given library).

3. Identifying Genes

50

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence

55   matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and and com-pared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

ILLUSTRATIVE APPLICATIONS

1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can by chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al*. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2 (1989).

3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, J. Clin. Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in:Handbook of Experimental Immunology, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0. 1 to 0. 2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in:Manual of Clinical Immunology, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of Staphylococcal disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunothereapeutic reagent.

3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS:1-5,191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-

mately the same. Th PCR primers and amplified DNA of this Example find use in the Examples that follow.

4. G ne expression from DNA S quenc s Corresponding to ORFs

A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.,* U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of theLTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

Alternativly and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.,* cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* ExpressTM Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	1419	757	emb|X17301|SAHD	S.aureus DNA for hld gene and for part of agr gene	100	663	663
1	2	3273	2452	emb|X53543|SAAG	S.aureus agrA, agrB and hld genes	99	809	822
1	5	6418	5651	dbj|D14711|STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	223	768
5	1	807	439	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	81	216	369
5	4	5031	3571	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	95	424	1461
10	1	86	904	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	98	715	819
16	5	5302	6246	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	94	251	945
16	6	6249	7091	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	99	843	843
16	7	7084	7584	gb|U35773|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	99	342	501
20	1	995	549	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	443	447
20	2	1011	841	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	91	137	171
20	3	2010	1798	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	110	213
20	4	5300	3825	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	948	1476
20	5	4788	4282	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	309	507
26	1	2	145	gb|U41072|	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	100	126	144
26	2	84	557	gb|U41072|	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	99	430	474
26	3	763	3531	emb|X74219|SAIL	S.aureus gene for isoleucyl-tRNA synthetase	99	2769	2769
29	3	1261	4392	gb|U66665|	Staphylococcus aureus DNA fragment with class II promoter activity	100	117	3132
31	14	14977	13463	emb|X73889|SAP1	S.aureus genes P1 and P2	99	1351	1515
31	15	14241	13855	emb|X73889|SAP1	S.aureus genes P1 and P2	98	258	387
38	17	14284	13112	gb|M12715|	S.aureus geh gene encoding lipase (glycerol ester hydrolase)	100	372	1173
38	19	13434	15518	gb|M12715|	S.aureus geh gene encoding lipase (glycerol ester hydrolase)	100	2085	2085

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
46	2	519	1727	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8D, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1209	1209
46	3	1720	2295	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	576	576
46	4	2259	3182	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	924	924
46	5	3173	4498	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1283	1326
46	6	4536	5720	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1185	1185
46	7	6455	6120	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	278	336
48	1	2	955	gb|L25893|	Staphylococcus aureus recA gene, complete cds	99	954	954
50	3	4465	2924	emb|X85029|SAAH	S.aureus AhpC gene	100	88	1542
50	4	4108	3515	emb|X85029|SAAH	S.aureus AhpC gene	98	540	594
54	3	5074	3392	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	1668	1683
54	4	4865	4122	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	99	720	744
54	5	5056	4562	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	463	495
54	6	11386	8300	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	100	3087	3087
58	3	1743	2819	emb|XR7104|SAPB	S.aureus mdr, pbp4 and tagD genes (SG511-55 isolate)	89	68	1077
58	4	2858	3280	emb|X91786|SAPB	S.aureus abcA, pbp4, and tagD genes	99	423	423
58	5	6005	4701	emb|X91786|SAPB	S.aureus abcA, pbp4, and tagD genes	99	1305	1305
58	6	5677	5378	gb|J29478|	Staphylococcus aureus ABC transporter-like protein AbcA (abcA) gene, partial cds	100	300	300
58	7	5086	6840	emb|X91786|SAPB	S.aureus abcA, pbp4, and tagD genes	99	1755	1755
72	1	888	445	gb|Z21854|	S.aureus agr gene encoding an accessory gene regulator protein, complete cds	100	444	444
72	2	2457	1453	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	99	673	1005

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nr. length
82	1	357	3917	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	2396	3561
82	2	4027	7677	emb|X89233|SARP	S.aureus DNA for rpoC gene	99	3171	3651
82	3	7745	8068	gb|U20869|	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	320	324
82	4	8103	8579	gb|U20869|	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	477	477
82	5	8618	8821	gb|U20869|	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	154	204
84	1	18	191	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	164	174
84	2	189	893	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	94	705	705
84	3	887	1660	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	774	774
84	4	1584	3503	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1920	1920
84	5	3394	4521	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1128	1128
84	6	4519	5643	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1125	1125
96	2	1245	3896	emb|Z18852|SACF	S.aureus gene for clumping factor	83	660	2652
97	2	625	882	gb|U41072|	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	97	68	258
111	1	3	452	gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	450	450
111	2	526	1041	gt|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	516	516
117	2	1278	1958	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	61	681

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
118	4	3787	4254	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	99	467	468
130	4	2597	3640	emb|X13290|SATN	Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing transposon Tn4003	78	956	1044
130	5	3813	4265	emb|216422|SADI	S.aureus dfrB gene for dihydrofolate reductase	98	416	453
130	6	4309	5172	emb|216422|SADI	S.aureus dfrB gene for dihydrofolate reductase	98	607	864
136	4	5296	6207	emb|X71437|SAGY	S.aureus genes gyrB, gyrA and recF (partial)	97	838	912
136	5	11680	8987	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	2694	2694
136	6	12886	10940	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	1947	1947
136	7	12592	11765	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	822	828
143	3	4171	2867	gb|U36379|	Staphylococcus aureus S-adenosylmethionine synthetase gene, complete cds	99	1305	1305
143	4	3100	4281	gb|L42943|	Staphylococcus aureus (clone KIN30) phosphoenolpyruvate carboxykinase (pcka) gene, complete cds	100	1170	1182
143	5	4254	4718	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pcka) gene, complete cds	100	449	465
143	9	6977	7261	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	75	285
143	10	9464	8161	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	1104	1104
143	11	11232	9748	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	1485	1485
143	12	10739	10320	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (mene), and o-succinylbenzoic acid synthetase (menc) genes, complete cds	100	332	420
152	5	2454	3437	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	305	984
152	6	3513	4820	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	1308	1308
152	7	4818	6230	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	1411	1413
153	1	387	1526	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	1140	1140
153	2	1877	2152	gb|S77055|	recF cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	100	276	276

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
153	3	2143	2289	gb|S77055|	recf cluster: dnaA=replisome assembly protein...gyrB=DNA gyrase beta subunit (Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt)	99	113	147
154	10	10792	9314	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	91	154	1479
154	11	9935	9615	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229	321
154	12	9943	10167	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	94	123	225
154	13	10069	11501	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	1326	1413
159	2	2195	1212	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	71	984
161	3	2596	2270	gb|H83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	92	203	327
162	1	1406	705	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	100	702	702
163	4	1263	1772	gb|U19770|	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	127	510
164	7	4774	9117	dbj|D86727|D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	99	3470	4344
168	7	7448	6447	gb|U21616|	Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds	100	1002	1002
168	8	9538	7961	gb|U21616|	Staphylococcus aureus cmp-binding-factor 1 (cbf1) and ORF X genes, complete cds	99	1158	1578
173	6	9240	7801	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	1440	1440
173	7	11252	9522	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	1731	1731
173	8	8285	8704	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	420	420
173	9	10168	9839	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	330	330
173	10	11815	10829	emb|X14827|SALA	Staphylococcus aureus lacC and lacD genes	100	987	987
173	11	12721	11774	emb|X14827|SALA	Staphylococcus aureus lacC and lacD genes	100	948	948
173	12	12838	12305	gb|H64724|	S.aureus tagatose 6-phosphate isomerase gene, complete cds	100	534	534
173	13	13243	12773	gb|H32103|	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	100	471	471
173	14	14633	13866	gb|H32103|	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	100	768	768

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
178	1	2	655	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	115	654
178	2	2201	1482	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	720	720
178	3	2361	1909	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	453	453
178	4	1551	1853	gb|U52961|	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	303	303
178	5	3541	2777	gb|L42945|	Staphylococcus aureus lytS and lytR genes, complete cds	99	765	765
178	6	3294	3025	gb|L42945|	Staphylococcus aureus lytS and lytR genes, complete cds	99	270	270
181	1	1114	590	gb|M61177|	S.aureus sigma factor (plaC) gene, complete cds	99	499	525
182	1	3	341	emb|X61307|SASP	Staphylococcus aureus spa gene for protein A	98	277	339
182	2	690	2312	gb|J01786|	S.aureus spa gene coding for protein A, complete cad	97	1332	1623
182	3	5861	4251	emb|X61307|SASP	Staphylococcus aureus spa gene for protein A	99	119	1611
185	1	3	824	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrosuinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	90	132	822
191	3	841	2760	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	99	1920	1920
191	4	2967	3143	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	99	177	177
191	5	5768	4566	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	99	250	1203
196	1	1741	872	gb|L36472|	Staphylococcus aureus lysyl-tRNA sythetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 16S ribosomal RNA (16S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene	99	870	870
198	3	1688	2011	emb|X92205|SAPT	S.aureus ptsH and ptsI genes	99	324	324
198	4	2005	2310	emb|X92205|SAPT	S.aureus ptsH and ptsI genes	97	304	306
202	1	163	1305	emb|X97985|SAI2	S.aureus orfs 1,2,3 & 4	99	1143	1143
202	2	1303	2175	emb|X73889|SAPI	S.aureus genes P1 and P2	94	444	873
210	1	3114	1558	dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1552	1557
210	2	2939	2232	gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	684	708
214	11	7429	7770	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	96	157	342

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
216	3	398	1318	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidins	88	265	921
219	2	1010	1073	dbj|D30690|STAN	Staphylococcus aureus genes for ORF35; HSP70; HSP40; ORF35, complete cds	100	60	738
219	3	2979	2035	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	99	945	945
219	4	4359	3196	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	99	1164	1164
219	5	7044	5176	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	98	1869	1869
219	6	6557	5883	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	99	675	675
219	7	6801	6334	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	98	468	468
221	8	10816	10034	gb|L19298|	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	91	67	783
223	1	2855	1506	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	102	1350
234	1	2	1357	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	100	176	1356
234	2	1694	2485	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	100	792	792
234	3	2648	3148	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	99	501	501
234	4	3120	4604	emb|X97985|SA12	S.aureus orfs 1,2,3 & 4	99	1305	1485
236	6	3826	5322	gb|U48826|	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	96	648	1497
248	1	2	403	emb|X62288|SAPE	S.aureus DNA for penicillin-binding protein 2	100	103	402
248	2	388	852	gb|L25426|	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	465	465
253	2	1539	1093	gb|U46541|	Staphylococcus aureus sarA gene, complete cds	96	447	447
254	2	150	1835	gb|U57060|	Staphylococcus aureus scdA gene, complete cds	94	142	1686
254	3	1973	2728	gb|U57060|	Staphylococcus aureus scdA gene, complete cds	99	756	756
260	1	2	1900	gb|M90693|	Staphylococcus aureus glycerol ester hydrolase (lip) gene, complete cds	99	1213	1899
265	1	1	942	dbj|D21121|STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	941	942

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
265	2	688	476	dbj|D21133|STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	211	213
265	3	2418	1765	dbj|D21133|STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	98	69	654
266	1	2	1018	dbj|D14711|STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	743	1017
282	1	1	525	gb|L72488|	hemB-porphobilinogen synthase (Staphylococcus aureus, SA1959, Genomic, 1087 nt)	100	110	525
282	2	516	1502	gb|S72488|	hemB-porphobilinogen synthase (Staphylococcus aureus, SA1959, Genomic, 1087 nt)	100	952	987
284	1	3	170	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	84	168
284	2	282	1034	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	712	753
284	3	1028	2026	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	979	999
284	4	1990	2202	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	187	213
289	3	1536	1991	gb|M32470|	S.aureus Sau3AI-restriction-enzyme and Sau3AI-modification-enzyme genes, complete cds	99	338	456
303	1	2	868	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	99	867	867
303	2	1409	2383	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	100	975	975
303	3	2367	3161	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hglC) genes, complete cds	99	793	795
305	1	2707	1355	dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1343	1353
311	1	2628	1315	gb|L42945|	Staphylococcus aureus lytS and lytR genes, complete cds	98	1314	1314
312	6	7019	7870	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	74	351	852
323	1	1998	1003	gb|U31175|	Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (ddh) gene, complete cds	98	996	996
326	1	1	237	emb|Y00356|SASP	Staphylococcus aureus V8 serine protease gene	100	108	237
338	1	687	388	emb|X64389|SALE	S.aureus leuF-PB3 gene for F component of leucocidin R	98	259	300
338	2	1828	1088	emb|X64389|SALE	S.aureus leuF-PR3 gene for F component of leucocidin R	97	137	741

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
342	2	579	1754	gb|U06462|	Staphylococcus aureus SA4 FtsZ (ftsZ) gene, complete cds	100	1176	1176
344	2	517	1248	emb|V01281|SANU	S.aureus mRNA for nuclease	98	732	732
349	1	457	230	gb|M20193|	S.aureus bacteriophage phi-11 attachment site (attB)	96	172	228
353	1	1016	516	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	187	501
353	2	1582	1046	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	99	537	537
356	1	3	674	gb|U20501|	Staphylococcus aureus MHC class II analog gene, complete cds	75	671	672
361	1	1	903	gb|L19298|	Staphylococcus aures phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	98	747	903
361	2	1103	1507	gb|L19298|	Staphylococcus aures phosphatidylinositol-specific phospholipase C (plc) gene, complete cds	97	68	405
373	1	3	1148	emb|X62288|SAPE	S.aureus DNA for penicillin-binding protein 2	99	1146	1146
389	3	1904	1248	emb|X62282|SATS	S.aureus target site DNA for IS431 insertion	97	349	657
400	1	1	540	emb|X61716|SAHL	S.aureus hlb gene encoding sphingomyelinase	99	389	540
400	2	1693	1187	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	178	507
408	1	1810	1049	gb|S76211|	aap23-alkaline shock protein 23 (methicillin resistant) (Staphylococcus aureus, 912, Genomic, 1160 nt)	99	163	762
418	1	2	217	gb|L41499|	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	216	216
418	2	854	639	db|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	100	188	216
421	2	1262	2509	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	1248	1248
422	1	2	325	gb|X02985|	S.aureus (strain RN450) transposon Tn554 insertion site	96	200	324
427	1	865	434	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	432	432
427	2	1829	1122	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	151	708
435	1	2	808	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	556	807
435	2	832	999	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	134	168
436	1	1341	685	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	97	657	657

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
436	2	2403	1657	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	100	294	747
442	1	347	1200	emb|X72700|SAPV	S.aureus genes for S and F components of Panton-Valentine leucocidin	84	204	954
445	2	1906	2178	gb|L01055|	Staphylococcus aureus gamma-hemolysin components A, B and C (hlgA, hlgB, hlgC) genes, complete cds	98	187	273
447	1	167	1078	gb|U19770|	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	100	51	912
447	2	1176	1784	gb|U19770|	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	597	609
454	3	7309	4319	emb|Z18852|SACF	S.aureus gene for clumping factor	75	653	2991
472	4	7896	5479	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	2418	2418
472	5	8120	6792	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	1328	1329
475	2	566	889	emb|X53543|SAAG	S.aureus agrA, agrB and hld genes	100	76	324
481	4	1922	1560	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothatical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	250	363
481	5	1244	1534	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothatical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	224	291
487	2	1388	1188	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	98	72	201
489	1	2737	1370	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	1368	1368
503	2	1135	653	gb|M83994|	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	108	483
511	3	1613	2242	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	323	630
511	4	3122	2700	gb|S76213|	asp23-alkaline shock protein 23 (methicillin resistant) [Staphylococcus aureus, 912, Genomic, 1360 nt]	96	423	423
520	2	758	1297	emb|X72014|SAFI	S.aureus fib gene for fibrinogen-binding protein	99	540	540
520	3	1436	1801	emb|X72013|SAFI	S.aureus fib gene for fibrinogen-binding protein	99	221	366
526	1	2150	1092	dbj|D17366|STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	641	1059

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
528	2	58	963	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	250	906
528	3	1098	2870	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	866	1773
530	1	3	434	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	99	432	432
530	2	1211	2395	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genus, complete cds, dehydroauinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	91	1185	1185
530	3	2409	2801	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroauinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	88	181	393
530	4	2690	3484	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	75	795
530	5	3482	4792	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimato-1-carboxyvinyltransforase (aroA) gene, complete cds; ORF3, complete cds	99	905	1311
530	6	4790	5380	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	196	591
539	1	3	338	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	99	336	336
539	2	336	527	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	100	189	192
554	1	727	165	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	54	363
554	2	2175	1252	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	918	924
554	3	1574	1374	gb|U73374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	96	122	201
584	2	1019	705	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	306	315
587	3	1475	4288	emb|Z18852|SACF	S.aureus gene for clumping factor	98	2588	2814
598	1	3881	1953	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	99	1871	1929

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
605	1	2	745	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	98	338	744
609	1	1628	816	emb|X76490|SAGL	S.aureus (bb270) glnA and glnR genes	100	495	813
614	1	1280	642	gb|M12103|	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	99	639	639
626	1	2508	1255	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	225,	1254
626	2	3315	2284	gb|M63176|	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	838	1032
629	1	1999	1001	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	990	999
629	2	1407	1195	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	98	194	213
631	2	5126	3228	emb|Z18852|SACF	S.aureus gene for clumping factor	82	489	1899
632	1	3	551	emb|Z10588|SAST	S.aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	549	549
631	2	529	1323	emb|Z10588|SAST	S.aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	795	795
651	1	1909	1070	gb|L19300|	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	478	840
657	2	1800	1105	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	456	696
662	1	908	456	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	369	453
662	2	230	475	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	246	246
662	3	746	1399	emb|X13404|SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	653	654
682	1	956	480	gb|U63177|	S.aureus sigma factor (plaC) gene, complete cds	100	136	477
685	1	1182	592	gb|U65000|	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	98	534	591
685	2	1716	1153	gb|U65000|	Staphylococcus aureus type-I signal peptidase SpsA (spsA) gene, and type-I signal peptidase SpsB (spsB) gene, complete cds	96	564	564
697	1	3	527	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	100	195	525
697	2	485	784	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	97	280	300

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
710	1	15	503	dbj|D86240|DR62	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	217	489
733	1	26	205	gb|M80252|	Staphylococcus aureus norAll99 gene (which mediates active efflux of fluoroquinolones), complete cds	97	140	180
741	1	1736	1197	dbj|D83951|STAL	Staphylococcus aureus DNA for LukH component, LukF-PV like component, complete cds	81	522	540
752	1	1	636	emb|Y00356|SASP	Staphylococcus aureus V8 serine protease gene	99	618	636
752	2	588	956	emb|Y00356|SASP	Staphylococcus aureus V8 serine protease gene	99	340	369
756	1	1108	709	emb|X01645|SATO	Staphylococcus aureus (Wood 46) gene for alpha-toxin	98	567	600
777	1	1582	950	emb|249245|SA42	S.aureus partial sod gene for superoxide dismutase	99	429	633
780	1	1111	557	gb|U20503|	Staphylococcus aureus MHC class-II analog gene, complete cds	86	550	555
784	1	73	687	gb|U63529|	Staphylococcus aureus novel antigen gene, complete cds	99	588	615
797	1	182	544	dbj|D14711|STAH	Staphylococcus aureus HSP10 and HSP60 genes	98	363	363
798	1	512	102	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	95	196	231
823	1	3	467	gb|S77055|	recF cluster: dnaA...replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, YB886, Genomic, 5 genes, 3573 nt]	99	156	465
848	1	348	175	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	174	174
848	2	476	318	gb|L25288|	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	100	131	159
866	1	792	397	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	395	396
883	1	1	285	dbj|D90119|STAN	S. aureus norA gene	99	131	285
884	1	606	334	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	98	265	273
884	2	716	522	emb|X52543|SAAG	S.aureus agrA, agrB and hld genes	100	195	195
912	2	517	681	emb|230588|SAST	S.aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	163	165
917	1	2	265	gb|N64724|	S.aureus tagatose 6-phosphate isomerase gene, complete cds	99	247	264
917	2	238	396	gb|N64724|	S.aureus tagatose 6-phosphate isomerase gene, complete cds	95	147	159
918	1	2426	1215	emt|X93205|SAPT	S.aureus ptsH and ptsI genes	99	1212	1212

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	IISP nt length	ORF nt length
967	1	1	411	dbj|D90119|STAN	S. aureus norA gene	97	395	411
991	1	672	337	emb|X52543|SAAG	S. aureus agrA, agrB and hld genes	99	336	336
1000	1	1117	845	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	78	190	273
1001	1	498	265	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	234	234
1010	1	1	285	gb|U21221|	Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds	99	224	285
1046	1	656	330	emb|X72700|SAPV	S. aureus genes for S and F components of Panton-Valentine leucocidins	85	205	327
1060	1	480	286	emb|X58434|SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	180	195
1073	1	1176	589	gb|X02985|	S. aureus (strain RN4501 transposon Tn554 insertion site	100	131	588
1079	1	3	230	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	228	228
1079	2	218	484	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	267	267
1079	1	460	645	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	186	186
1092	1	289	146	emb|X58434|SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	124	144
1143	1	1	243	gb|M63177|	S. aureus sigma factor (plaC) gene, complete cds	99	243	243
1157	1	2	136	gb|248003|SADN	S. aureus gene for DNA polymerase III	97	127	135
1189	1	720	361	gb|S74031|	norA=NorA (ISP794) [Staphylococcus aureus, NCTC 8325, insertion, 1820 nt]	99	360	360
1190	1	2	283	gb|M21854|	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	282	282
1190	2	1127	888	emb|X52543|SAAG	S. aureus agrA, agrB and hld genes	100	240	240
1225	1	2	163	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	97	124	162
1243	1	2	529	dbj|D86240|D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	495	528
1244	1	1	210	gb|S74031|	norA=NorA (ISP794) [Staphylococcus aureus, NCTC 8325, insertion, 1820 nt]	100	210	210
1301	1	41	472	emb|X76490|SAGL	S. aureus (bb270) glnA and glnR genes	99	299	432

TABLE 1

S. aureus - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1315	1	18	326	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	277	309
1519	1	2	175	dbj|D28879|STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	98	139	174
1663	1	1346	675	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	98	672	672
1797	1	644	324	gb|U71374|	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	321	321
1857	1	1	192	gb|M90536|	Staphylococcus aureus alpha-hemolysin gene, 3' end	98	192	192
1921	1	2	181	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	100	180	180
1957	1	2	346	gb|U60589|	Staphylococcus aureus novel antigen gene, complete cds	99	345	345
1988	1	1	402	dbj|D86240|D862	Staphylococcus aureus gene for unkown function and dlt operon dltA, dltB, dltC and dltD genes,complete cds	100	402	402
2100	1	414	208	gb|M63177|	S.aureus sigma factor (plaC) gene, complete cds	99	207	207
2109	1	1	402	gb|U66664|	Staphylococcus aureus DNA (fragment with class II promoter activity	99	131	402
2537	1	308	156	emb|X17688|SAFE	S.aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	153	153
2891	1	2	400	gb|L25426|	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	399	399
2950	1	778	398	dbj|D30690|STAN	Staphylococcus aureus genes for ORF17; HSP20; HSP40; HSP70; HSP90; ORF35; complete cds	100	358	381
2971	1	3	398	gb|U51132|	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (menE), and o-succinylbenzoic acid synthetase (menC) genes, complete cds	97	272	396
2978	1	618	328	gb|U31979|	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinate synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	98	250	291
2985	1	832	464	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	98	347	369
3006	1	2170	1784	gb|U11779|	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRNV30 16S-23S rRNA spacer region	87	82	387
3008	1	473	228	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35; complete cds	88	178	237
3008	2	451	281	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35; complete cds	97	120	171

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP length	ORF nt length
3011	1	793	398	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	93	72	396
3019	1	2	235	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	234	234
3023	1	81	233	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	87	100	153
3029	2	90	287	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	135	198
3039	1	18	164	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	97	135	147
3039	2	70	327	gb|U51133|	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	77	183	258
3056	1	3	215	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	213	213
3059	1	1	261	dbj|D30690|STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	234	261
3073	1	27	284	gb|U06451|	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229	258
3074	1	2	397	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	96	250	396
3088	1	3	239	dbj|D86727|DB67	Staphylococcus aureus DNA for DNA polymerase III, complete cds	95	215	237
3097	1	444	244	emu|748003|SAHH	S.aureus gene for DNA polymerase III	97	160	201
3102	1	307	155	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	142	153
3121	1	568	398	emb|X58434|SAPD	S.aureus pdhA, pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	100	88	171
3125	1	463	233	emb|A89213|SARP	S.aureus DNA for rpoC gene	98	192	231
3133	1	2	175	emb|218852|SACF	S.aureus gene for clumping factor	96	154	174
3160	1	420	211	dbj|D10489|STAG	Staphylococcus aureus gene for DNA gyrase A and B, complete cds	89	197	210
3176	1	1	378	emb|X58434|SAPD	S.aureus pdhA, pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	91	378
3192	1	420	211	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	72	210
3210	1	3	143	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	96	141	141

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
3232	3	2106	1202	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	71	257	825
3538	1	?	304	emb|X89233|SARP	S.aureus DNA for rpoC gene	99	356	393
3543	1	392	634	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	99	102	243
3555	1	637	320	emb|Z18852|SACF	S.aureus gene for clumping factor	99	307	318
3559	1	3	182	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	100	141	180
3559	2	95	313	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	98	174	219
3563	1	278	141	gb|U35777|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	100	79	138
3563	2	527	363	gb|U35777|	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	98	162	165
3566	1	3	422	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	98	175	420
3588	1	2	262	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	253	261
3591	1	1	350	gb|U03477|	S.aureus enzyme IIf-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	145	348
3600	1	758	381	emb|Z18852|SACF	S.aureus gene for clumping factor	72	246	378
3602	1	788	396	emb|Z18852|SACF	S.aureus gene for clumping factor	98	319	393
3656	1	1013	528	emb|Z18852|SACF	S.aureus gene for clumping factor	84	403	486
3682	1	3	236	emb|X64172|SARP	S.aureus rplB, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	231	234
3682	2	224	415	emb|X64172|SARP	S.aureus rplB, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	112	192
3693	1	758	423	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	229	336
3702	1	593	354	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	55	81	240
3725	1	924	463	emb|Z18852|SACF	S.aureus gene for clumping factor	71	367	462
3761	1	809	450	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	85	333	360
3767	1	1	402	emb|X64172|SARP	S.aureus rplB, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	387	402

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
3775	1	2	286	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	227	285
3786	1	456	229	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	204	228
3786	2	542	366	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	95	123	177
379R	1	3	251	emb|X17679|SACO	Staphylococcus aureus coa gene for coagulase	99	249	249
3813	1	793	398	gb|J04155|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	396	396
3819	1	184	402	emb|X68425|SA23	S.aureus gene for 23S rRNA	99	161	219
3844	1	932	468	gb|U48826|	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	87	204	465
3845	1	1	381	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	94	356	381
3856	1	798	400	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecRI gene and unknown ORF, complete cds	76	192	399
3859	1	1049	573	emb|Z18852|SACF	S.aureus gene for clumping factor	85	307	477
3871	1	650	327	gb|M76714|	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	299	324
3876	1	2	253	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	217	252
3877	1	572	288	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	209	285
3878	1	1	237	emb|X58434|SAPD	S.aureus pdhA, pdhC and pdhD genes for pyruvate docarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	96	155	237
3888	1	3	173	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	98	171	171
3893	1	1	183	emb|X89233|SARP	S.aureus DNA for rpoC gene	100	170	183
3893	2	181	357	emb|X89233|SARP	S.aureus DNA for rpoC gene	98	79	177
3894	1	3	485	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	450	483
3895	1	836	420	gb|J04155|	S.aureus (fibronectin-binding protein (fnbA) mRNA, complete cds	99	411	417
3905	1	48	239	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	159	192
3905	2	188	400	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	97	88	213

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	Percent ident	HSP nt length	ORF nt length
3910	1	3	359	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	278	357
3915	1	1	330	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	175	330
3964	1	691	347	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	100	295	345
4007	1	199	390	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	98	163	192
4036	1	3	371	dbj|D10489|STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	339	369
4046	1	692	348	emb|Z18852|SACF	S.aureus gene for clumping factor	87	221	345
4060	1	1	375	emb|Z18852|SACF	S.aureus gene for clumping factor	96	271	375
4061	1	860	432	emb|Z48003|SADN	S.aureus gene for DNA polymerase III	99	429	429
4062	1	606	304	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	198	303
4085	1	58	402	gb|U11786|	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRUV42 16S-23S rRNA spacer region	98	127	345
4088	1	2	301	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	227	300
4093	1	2	277	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	99	276	276
4097	1	1	402	emb|Z18852|SACF	S.aureus gene for clumping factor	74	307	402
4116	1	22	402	gb|L05004|	Staphylococcus aureus dehydroquinate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds: ORF3, complete cds	98	157	381
4125	1	240	401	gb|U73374|	Staphylococcus aureus type 8 capsule genes, capBA, capBB, capBC, capBD, capBE, capBF, capBG, capBH, capBI, capBJ, capBK, capBL, capBM, capBN, capBO, capBP, complete cds	100	86	162
4149	1	35	247	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	200	213
4151	1	629	366	gb|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	87	150	264
4154	1	754	398	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	297	357
4179	1	1	294	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	240	294

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt Length
4203	1	1	255	emb|X89233|SARP	S.aureus DNA for rpoC gene	99	239	255
4206	1	1	303	emb|Z18852|SACF	S.aureus gene for clumping factor	100	236	303
4206	2	195	344	emb|Z18852|SACF	S.aureus gene for clumping factor	95	65	150
4208	1	108	314	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	89	76	207
4216	1	656	330	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	326	327
4226	1	594	298	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	97	132	297
4260	1	216	383	gb|U11784|	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRNV40 16S-23S rRNA spacer region	83	141	168
4272	1	355	179	emb|Z48003|SAUN	S.aureus gene for DNA polymerase III	100	164	177
4276	1	4	177	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	99	150	174
4277	1	1	270	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	265	270
4282	1	691	377	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	282	315
4291	1	179	191	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	183	189
4295	1	3	329	emb|X16457|SAST	Staphylococcus aureus gene for staphylocoagulase	94	144	327
4313	1	435	280	gb|L11530|	Staphylococcus aureus transfer RNA sequence with two rRNAs	100	94	156
4315	1	3	185	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	158	183
4315	2	101	310	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	75	210
4327	1	1	294	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	294	294
4360	1	601	319	gb|U02910|	Staphylococcus aureus ATCC 25923 16S rRNA gene, partial sequence	100	116	285
4364	1	3	146	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	95	140	144
4388	1	167	310	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	73	119	144

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
4401	1	2	313	emb|X62992|SAFN	S.aureus fnbB gene for fibronectin binding protein B	97	243	312
4421	1	36	281	dbj|D12572|STA2	Staphylococcus aureus rrnA gene for 23S ribosomal RNA	100	112	246
4426	1	3	293	emb|218852|SACF	S.aureus gene for clumping factor	85	185	291
4428	1	493	248	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	139	246
4462	1	2	271	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	270	270
4466	1	1	240	emb|218852|SACF	S.aureus gene for clumping factor	99	231	240
4469	1	1	312	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	265	312
4485	1	3	263	gb|L43098|	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	259	261
4492	1	74	400	gb|M86227|	Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit (gyrA) gene, complete cds	85	104	127
4497	1	515	269	emb|218852|SACF	S.aureus gene for clumping factor	99	213	267
4529	1	2	172	emb|X64172|SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	151	171
4547	1	1	300	emb|X62092|SAFN	S.aureus fnbD gene for fibronectin binding protein D	100	157	300
4554	1	318	160	emb|218852|SACF	S.aureus gene for clumping factor	84	126	159
4565	1	9	227	emb|218852|SACF	S.aureus gene for clumping factor	84	213	219
4569	1	79	222	emb|218852|SACF	S.aureus gene for clumping factor	98	127	144
4608	1	22	216	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	92	168	195
4614	1	464	234	emb|218852|SACF	S.aureus gene for clumping factor	86	169	231
4623	1	105	302	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	152	198
4632	1	18	206	gb|J03479|	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	183	189
4646	1	1	222	emb|218852|SACF	S.aureus gene for clumping factor	84	100	222
4687	1	2	166	gb|J04151|	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	156	165

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	percent ident	HSP nt length	ORF nt length
4695	1	313	158	gi|L14017|	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	155	156
4703	1	1	153	emb|X58434|SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	103	153

TABLE 1

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
20	6	5089	4679	gi|511839	ORF1 [Staphylococcus bacteriophage phi 11]	100	100	411
149	3	2032	1577	pir|B49703|B497	int gene activator RinA - bacteriophage phi 11	100	100	456
149	5	2109	1912	gi|166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	100	100	198
349	2	558	409	gi|166159	integrase (int) [Staphylococcus bacteriophage phi 11]	100	100	150
398	1	1372	707	gi|166159	Integrase (int) [Staphylococcus bacteriophage phi 11]	100	99	666
398	2	783	1001	gi|455128	excisionase (xis) [Staphylococcus bacteriophage phi 11]	100	100	219
502	4	1914	1744	gi|1204912	H. influenzae predicted coding region HI0660 [Haemophilus influenzae]	100	71	171
849	1	2	262	gi|1373002	polyprotein [Bean common mosaic virus]	100	46	261
1349	1	277	140	gi|143359	protein synthesis initiation factor 2 (infB) [Bacillus subtilis] gi|49319 IF2 gene product [Bacillus subtilis]	100	82	138
2880	1	21	308	gi|862933	protein kinase C inhibitor-I [Homo sapiens]	100	98	288
3045	1	428	216	gi|1354211	PET112-like protein [Bacillus subtilis]	100	100	213
4168	2	571	398	gi|1354211	PET112-like protein [Bacillus subtilis]	100	100	174
331	1	2	247	gi|426473	nusG gene product [Staphylococcus carnosus]	98	95	246
207	2	1272	1463	gi|460259	enolase [Bacillus subtilis]	97	90	192
331	2	395	H50	gi|581638	hit protein [Staphylococcus carnosus]	97	?3	456
366	1	39	215	gi|66161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	97	95	177
680	3	718	936	gi|426473	nusG gene product [Staphylococcus carnosus]	97	97	219
1578	1	284	144	gi|1319950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	97	79	141
157	1	321	518	gi|1022726	unknown [Staphylococcus haemolyticus]	96	88	198
205	13	16470	16147	gi|1165302	S10 [Bacillus subtilis]	96	91	324
3919	1	48	401	gi|971784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	96	61	354
4133	1	830	417	gi|1022726	unknown [Staphylococcus haemolyticus]	96	84	414
4168	1	708	355	gi|1354211	PET112-like protein [Bacillus subtilis]	96	95	354
4207	1	312	157	gi|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC 1.5.99.7) - Mycoplasma capricolum (SQC3) (fragment)	96	86	156

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4227	2	152	331	gi|871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	96	81	180
4416	1	570	286	gi|1022726	unknown [Staphylococcus haemolyticus]	96	84	285
22	1	858	430	gi|511070	UreD [Staphylococcus xylosus]	95	88	429
22	7	4362	4036	gi|581787	urease gamma subunit [Staphylococcus xylosus]	95	79	327
82	6	8794	9114	pir|JG0008|JG00	ribosomal protein-S7 - Bacillus stearothermophilus	95	83[1]	321
154	9	9280	7838	gi|1354211	PEPT112-like protein [Bacillus subtilis]	95	92	1443
186	3	2798	2055	gi|1514656	serine O-acetyltransferase [Staphylococcus xylosus]	95	87	744
205	5	4406	4014	gi|142462	ribosomal protein S11 [Bacillus subtilis]	95	85	393
205	7	5017	4793	gi|142459	initiation factor 1 [Bacillus subtilis]	95	84	225
205	21	11365	10991	gi|1044974	ribosomal protein L14 [Bacillus subtilis]	95	93	375
259	5	7288	6644	sp|P47995|YSEA_	HYPOTHETICAL PROTEIN IN SECA 5'REGION (ORF1) (FRAGMENT).	95	85	645
302	3	795	1097	gi|40186	homologous to E.coli ribosomal protein L27 [Bacillus subtilis] 1|143592 L27 ribosomal protein [Bacillus subtilis] 1r|C21895|C21895 ribosomal protein L27 - Bacillus subtilis p|PU5657|RL27_BACSU 50S RIBOSOMAL PROTEIN L27 [BL30] [BL24], 1|40175 L24 gene prod	95	89	303
310	1	579	1523	gi|1177684	chorismate mutase [Staphylococcus xylosus]	95	92	945
414	1	2	163	pir|C48396|C483	ribosomal protein L34 - Bacillus stearothermophilus	95	90	162
4185	2	125	277	gi|1276841	glutamate synthase (GOGAT) [Porphyra purpurea]	95	86	153
22	2	1028	723	gi|511069	UreF [Staphylococcus xylosus]	94	91	306
22	5	5046	3310	gi|410516	urease alpha subunit [Staphylococcus xylosus]	94	85	1737
60	4	815	1372	gi|666116	glucose kinase [Staphylococcus xylosus]	94	87	558
205	18	10012	9516	gi|1044978	ribosomal protein S8 [Bacillus subtilis]	94	78	477
326	4	3178	2542	gi|557492	dihydroxynapthoic acid (DHNA) synthetase [Bacillus subtilis] gi|143186 dihydroxynapthoic acid (DHNA) synthetase [Bacillus ubtilis]	94	85	837
414	3	737	955	gi|467386	thiophen and furan oxidation [Bacillus subtilis]	94	77	219
426	3	2260	1823	gi|1263908	putative [Staphylococcus epidermidis]	94	87	438
534	1	2	355	gi|633650	enzyme II(mannitol) [Staphylococcus carnosus]	94	84	354
1017	1	2	229	gi|149435	putative [Lactococcus lactis]	94	73	228
3098	1	330	184	gi|413952	lpa-28d gene product [Bacillus subtilis]	94	50	147

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3232	1	630	316	gi|1022725	unknown [Staphylococcus haemolyticus]	94	84	315
42	5	2089	2259	pir|B48196|B483	ribosomal protein L33 - Bacillus stearothermophilus	93	81	171
101	2	1745	1383	gi|155345	arsenic efflux pump protein [Plasmid pSX267]	93	82	363
205	24	12227	11865	sp|P14577|RL16_	50S RIBOSOMAL PROTEIN L16	93	83	363
259	4	8291	5673	gi|499335	secA protein [Staphylococcus carnosus]	93	85	2619
275	1	2226	1114	gi|613650	enzyme II(mannitol) [Staphylococcus carnosus]	93	86	1113
444	6	6207	5773	gi|1022726	unknown [Staphylococcus haemolyticus]	93	81	435
491	1	152	622	gi|46912	ribosomal protein L13 [Staphylococcus carnosus]	93	88	471
607	6	1674	2033	gi|1022726	unknown [Staphylococcus haemolyticus]	93	83	360
653	1	973	488	gi|580890	translation initiation factor IF3 [AA 1-172] [Bacillus stearothermophilus]	93	77	486
1864	1	3	194	gi|306553	ribosomal protein small subunit [Homo sapiens]	93	93	192
2997	1	28	300	gi|143390	carbamyl phosphate synthetase [Bacillus subtilis]	93	82	273
3232	2	907	596	gi|1022725	unknown [Staphylococcus haemolyticus]	93	84	312
3761	2	794	621	gi|1022725	unknown [Staphylococcus haemolyticus]	93	88	174
16	1	3	374	gi|142781	putative cytoplasmic protein; putative [Bacillus subtilis] sp|P37954|UVRB_BACSU EXCINUCLEASE ABC SUBUNIT B (DINA PROTEIN) FRAGMENT).	92	83	372
11	7	5915	6124	gi|1136430	KIAA0185 protein [Homo sapiens]	92	46	210
56	19	26483	27391	gi|467401	unknown [Bacillus subtilis]	92	80	909
69	6	5882	6130	gi|530200	trophoblastin [Ovis aries]	92	53	249
145	3	2568	2038	gi|1022725	unknown [Staphylococcus haemolyticus]	92	80	531
171	3	2760	2362	gi|517475	D-amino acid transaminase [Staphylococcus haemolyticus]	92	86	399
205	12	7495	6962	gi|49189	secY gene product [Staphylococcus carnosus]	92	85	534
205	19	10812	10255	gi|1044976	ribosomal protein L5 [Bacillus subtilis]	92	82	558
219	1	710	357	gi|1303812	VqeV [Bacillus subtilis]	92	88	354
144	3	1575	1805	gi|1405474	cspC protein [Bacillus cereus]	92	85	231
699	1	20	361	gi|413999	lpa-75d gene product [Bacillus subtilis]	92	81	342
1343	1	2	160	pir|A45434|A454	ribosomal protein L19 - Bacillus stearothermophilus	92	84	159

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1958	1	524	264	gi|407908	Eiiscr [Staphylococcus xylosus]	92	80	261
3578	2	718	386	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	92	78	333
3585	1	644	324	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	92	81	321
3640	1	4	402	gi|1022726	unknown [Staphylococcus haemolyticus]	92	81	399
4362	1	14	178	gi|450688	hadH gene of EcoprI gene product [Escherichia coli] pir|S38437|S38437 hadH protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	92	78	165
4446	1	358	182	gi|1022725	unknown [Staphylococcus haemolyticus]	92	82	177
4549	1	462	232	gi|1022726	unknown [Staphylococcus haemolyticus]	92	80	231
4626	1	3	224	gi|1022725	unknown [Staphylococcus haemolyticus]	92	84	222
2	4	3980	4531	gi|535349	CodW [Bacillus subtilis]	91	74	552
28	1	2	1126	gi|1001376	hypothetical protein [Synechocystis sp.]	91	78	1125
60	5	1354	1701	gi|226044	orf2 downstream of glucose kinase [Staphylococcus xylosus]	91	80	348
101	1	1989	1036	gi|150728	arsenic efflux pump protein [Plasmid pI258]	91	80	954
187	2	412	1194	gi|142559	ATP synthase alpha subunit [Bacillus megaterium]	91	79	783
205	22	11579	11298	gi|40149	S17 protein (AA 1-87) [Bacillus subtilis]	91	83	282
206	7	8184	10262	gi|1072418	glcA gene product [Staphylococcus carnosus]	91	83	2079
306	2	3885	2326	gi|143012	GMP synthetase [Bacillus subtilis]	91	78	1560
306	3	5319	3826	gi|467399	IMP dehydrogenase [Bacillus subtilis]	91	79	1494
310	3	2194	3207	gi|1177685	ccpA gene product [Staphylococcus xylosus]	91	81	1014
343	4	2974	3150	gi|949974	sucrose repressor [Staphylococcus xylosus]	91	82	177
480	3	1606	3042	gi|433991	ATP synthase subunit beta [Bacillus subtilis]	91	85	1437
536	3	2026	1280	gi|143366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir|C29326|W28SDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	91	79	747
552	1	1064	615	gi|297874	fructose-bisphosphate aldolase [Staphylococcus carnosus] pir|A49943|A49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TM300)	91	79	450
637	1	1	1536	gi|143597	CTP synthetase [Bacillus subtilis]	91	79	1536
859	1	21	359	gi|385178	unknown [Bacillus subtilis]	91	66	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1327	1	339	510	gi|496558	orfX (Bacillus subtilis)	91	71	192
2515	1	466	275	gi|511070	UreG (Staphylococcus xylosus)	91	85	192
2594	1	2	202	gi|146824	beta-cystathionase (Escherichia coli)	91	75	201
1764	1	847	423	gi|1022725	unknown (Staphylococcus haemolyticus)	91	78	423
4011	1	127	495	gi|1022726	unknown (Staphylococcus haemolyticus)	91	79	369
4227	1	1	177	gi|295464	ATPase (Lactococcus lactis)	91	66	177
42	3	815	1033	gi|520401	catalase (Haemophilus influenzae)	90	86	219
51	8	3717	4607	gi|580899	OppF gene product (Bacillus subtilis)	90	74	891
129	3	5317	4001	gi|1146206	glutamate dehydrogenase (Bacillus subtilis)	90	76	1317
164	17	16628	16933	sp|P05766|RS15_	30S RIBOSOMAL PROTEIN S15 (BS18).	90	74	306
171	5	2983	2819	gi|517475	D-amino acid transaminase (Staphylococcus haemolyticus)	90	78	165
205	4	4497	3550	gi|142463	RNA polymerase alpha-core-subunit (Bacillus subtilis)	90	76	948
205	6	4748	4410	gi|1044989	ribosomal protein S13 (Bacillus subtilis)	90	73	339
205	10	7165	6404	gi|49189	secY gene product (Staphylococcus carnosus)	90	81	762
205	11	6645	6472	gi|49189	secY gene product (Staphylococcus carnosus)	90	78	174
205	27	13692	13345	gi|786157	Ribosomal Protein S19 (Bacillus subtilis)	90	79	348
205	31	15858	15496	gi|1165303	L3 (Bacillus subtilis)	90	79	363
260	5	7023	5773	gi|1161380	IcaA (Staphylococcus epidermidis)	90	78	1251
299	6	3378	3947	gi|467440	'phosphoribosylpyrophosphate synthetase (Bacillus subtilis) gi|40218 PRPP synthetase (AA 1-317) (Bacillus subtilis)	90	78	570
320	2	1025	1717	gi|312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) (Bacillus aldolyticus)	90	75	693
330	4	1581	1769	gi|986963	beta-tubulin (Sporidiobolus pararoseus)	90	80	189
369	1	954	523	pir|S34762|S347	L-serine dehydratase beta chain - Clostridium sp.	90	77	432
557	1	3	188	gi|1511589	M. jannaschii predicted coding region MJ1624 (Methanococcus jannaschii)	90	54	186
663	2	667	1200	gi|143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) pir|JT0481|YWDS tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus ubtilis	90	73	534
717	1	1	261	gi|143065	hubat (Bacillus stearothermophilus)	90	79	261
745	4	1059	865	gi|1205433	H. influenzae predicted coding region HI1190 (Haemophilus influenzae)	90	81	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1007	1	386	565	gi|143366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir|C29326|WZBSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	90	77	180
1054	1	579	331	gi|1033122	ORF_f729 [Escherichia coli]	90	50	249
1156	1	117	707	gi|1477776	ClpP [Bacillus subtilis]	90	80	591
1180	1	408	205	gi|1377831	unknown [Bacillus subtilis]	90	74	204
1253	1	1	462	gi|40046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] lr|S15936|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	90	75	462
2951	1	3	269	gi|144816	formyltetrahydrofolate synthetase (FTHFS) (ttg start codon) (EC .3.4.3) [Moorella thermoacetica]	90	76	267
3140	1	227	166	gi|1070014	protein-dependent [Bacillus subtilis]	90	52	162
4594	1	3	233	gi|871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	90	76	231
87	1	1028	1750	gi|467327	unknown [Bacillus subtilis]	89	75	723
112	1	2	505	gi|153741	ATP-binding protein [Streptococcus mutans]	89	77	504
118	1	120	398	gi|1303804	YqeQ [Bacillus subtilis]	89	75	279
128	4	3545	3757	gi|460257	triose phosphate isomerase [Bacillus subtilis]	89	84	213
164	12	11667	12755	gi|139954	IP2 (aa 1-741) [Bacillus stearothermophilus]	89	80	1089
205	11	7875	7405	ni|216338	ORF for L15 ribosomal protein [Bacillus subtilis]	89	76	471
205	32	16152	15823	gi|1165303	L3 [Bacillus subtilis]	89	80	330
270	3	2407	2207	pir|C41902|C419	arsenate reductase (EC 1.-.-.-) - Staphylococcus xylosus plasmid pSX267	89	81	201
395	2	157	672	gi|520574	glutamate racemase [Staphylococcus haemolyticus]	89	80	516
494	1	3	839	gi|396259	protease [Staphylococcus epidermidis]	89	77	837
510	1	1	444	gi|40046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] lr|S15936|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	89	74	444
615	1	2124	1210	gi|1303812	YqeV [Bacillus subtilis]	89	74	915
841	1	18	341	gi|1165303	L3 [Bacillus subtilis]	89	80	324
1111	1	352	813	gi|47146	thermonuclease [Staphylococcus intermedius]	89	70	462
1875	1	2	256	gi|1205108	ATP-dependent protease binding subunit [Haemophilus influenzae]	89	82	255
2963	1	11	367	gi|467458	cell division protein [Bacillus subtilis]	89	83	157

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3020	1	90	362	gi|1239988	hypothetical protein [Bacillus subtilis]	89	66	273
3565	1	2	400	gi|1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	89	75	399
3586	1	105	314	gi|580832	ATP synthase subunit gamma [Bacillus subtilis]	89	82	210
3629	1	794	399	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	89	78	396
368R	1	2	400	gi|1146206	glutamate dehydrogenase [Bacillus subtilis]	89	75	399
3699	1	794	399	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	89	75	396
4016	1	428	216	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	89	71	213
4177	1	471	301	gi|149926	putative [Lactococcus lactis]	89	76	171
4436	1	601	302	gi|1022725	unknown [Staphylococcus haemolyticus]	89	80	300
4635	1	320	162	gi|1022725	unknown [Staphylococcus haemolyticus]	89	73	159
2	2	1330	2676	gi|520754	putative [Bacillus subtilis]	88	76	1347
42	2	468	848	sp|P42321|CATA_	CATALASE (EC 1.11.6)	88	76	381
53	5	6389	4722	gi|474177	alpha-D-1,4-glucosidase [Staphylococcus xylosus]	88	80	1668
56	16	18018	18617	gi|467611	recombination protein [Bacillus subtilis]	88	77	600
60	3	376	843	gi|666116	glucose kinase [Staphylococcus xylosus]	88	77	468
70	2	1583	1245	gi|44095	replication initiator protein [Listeria monocytogenes]	88	74	339
82	8	11514	12719	pir|A60663|A606	translation elongation factor Tu - Bacillus subtilis	88	79	1206
101	7	4179	4391	gi|167181	serine/threonine kinase receptor [Brassica napus]	88	77	213
114	8	7732	8232	gi|1022726	unknown [Staphylococcus haemolyticus]	88	72	501
118	2	308	2011	gi|1303804	YqeQ [Bacillus subtilis]	88	77	1704
141	3	657	1136	gi|1405446	transketolase [Bacillus subtilis]	88	72	480
148	7	5871	6116	gi|1118002	dihydropteroate synthase [Staphylococcus haemolyticus]	88	78	246
165	3	1428	2231	gi|40053	phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis] tr|S11730|YFDSA phenylalanine--tRNA ligase (EC 6.1.1.20) alpha sin - Bacillus subtilis	88	80	804
205	28	15027	14185	gi|1165306	L2 [Bacillus subtilis]	88	82	843
223	1	1569	898	gi|1303840	YqfS [Bacillus subtilis]	88	78	672
235	1	2	1975	gi|452309	valyl-tRNA synthetase [Bacillus subtilis]	88	76	1974

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
339	3	2060	1566	gi|1118002	dihydropteroate synthase (Staphylococcus haemolyticus)	88	73	495
443	4	4325	2928	gi|558559	pyrimidine nucleoside phosphorylase [Bacillus subtilis]	88	73	1398
532	1	3	419	gi|143797	valyl-tRNA synthetase [Bacillus stearothermophilus] sp|P19931|SYV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE-TRNA LIGASE) (VALRS).	88	78	417
534	3	2504	2988	gi|153049	mannitol-specific enzyme-III [Staphylococcus carnosus] pir|JQ0088|JQ0088 phosphotransferase system enzyme II (EC .7.1.69), mannitol-specific, factor III - Staphylococcus carnosus sp|P17876|PTMA_STACA PTS SYSTEM, MANNITOL-SPECIFIC IIA COMPONENT EIIA-MTL) (	88	82	465
705	2	584	399	gi|710018	nitrite reductase (nirB) [Bacillus subtilis]	88	70	186
1000	2	1424	1309	gi|1022726	unknown [Staphylococcus haemolyticus]	88	78	516
1299	1	587	324	gi|401786	phosphomannomutase [Mycoplasma pirum]	88	55	264
1341	2	170	400	gi|39963	ribosomal protein L20 (AA 1-119) [Bacillus stearothermophilus] lr|S05348|R35B520 ribosomal protein L20 - Bacillus aerothermophilus	88	82	231
1386	1	41	214	pir|B47154|B471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	88	71	174
1386	2	183	533	pir|D47154|B471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	88	73	351
2949	1	704	399	gi|535350	CodX [Bacillus subtilis]	88	73	306
2984	1	5	169	gi|218277	O-acetylserine(thiol) lyase [Spinacia oleracea]	88	70	165
3015	1	1	138	gi|493083	dihydroxyacetone kinase [Citrobacter freundii]	88	67	138
3089	1	3	152	gi|606055	ORF_F746 [Escherichia coli]	88	88	150
3917	1	817	410	gi|143378	pyruvate decarboxylase (E-1) beta subunit [Bacillus subtilis] gi|1377836 pyruvate decarboxylase E-1 beta subunit [Bacillus subtilis]	88	77	408
4199	1	680	342	gi|1515938	aconitase [Bacillus subtilis]	88	82	339
4201	1	734	369	gi|1515938	glutamate synthase (ferredoxin) [Synechocystis sp.] pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	366
4274	1	1	336	gi|1515938	glutamate synthase (ferredoxin) [Synechocystis sp.] pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	336
4308	1	794	399	gi|1146206	glutamate dehydrogenase [Bacillus subtilis]	88	71	396
2	5	4570	6000	gi|535350	CodX [Bacillus subtilis]	87	70	1431
52	8	6781	6482	gi|1064791	function unknown [Bacillus subtilis]	87	66	300

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
73	3	1584	2480	gi|142992	glycerol kinase (glpK) (EC 2.7.1.30) [Bacillus subtilis] pir|B45868|B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp|P18157|GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	87	72	897
98	12	8813	9100	gi|467433	unknown [Bacillus subtilis]	87	62	288
124	4	4265	2988	gi|556886	serine hydroxymethyltransferase [Bacillus subtilis] pir|S49363|S49363 serine hydroxymethyltransferase - Bacillus ubtilis	87	77	1278
124	6	4457	4032	gi|556883	Unknown [Bacillus subtilis]	87	66	426
148	5	3741	4559	gi|467460	unknown [Bacillus subtilis]	87	70	819
164	13	12710	13810	gi|39954	IF2 (aa 1-741) [Bacillus stearothermophilus]	87	72	1101
177	2	1104	2126	gi|467385	unknown [Bacillus subtilis]	87	78	1023
199	1	1982	1158	gi|143527	iron-sulfur protein [Bacillus subtilis]	87	77	825
199	2	4717	2933	pir|A27763|A277	succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis	87	80	1785
205	23	11782	11543	gi|1044972	ribosomal protein L29 [Bacillus subtilis]	87	78	240
205	25	13275	12607	gi|1165309	S3 [Bacillus subtilis]	87	75	669
222	1	2033	1107	gi|1177249	rec211 gene product [Bacillus subtilis]	87	70	927
236	3	1635	1333	gi|1146198	[erredoxin [Bacillus subtilis]	87	80	303
246	5	2585	2292	gi|467373	ribosomal protein S18 [Bacillus subtilis]	87	77	294
260	2	4189	3422	gi|1161382	IcaC [Staphylococcus epidermidis]	87	72	768
320	3	1696	2391	gi|312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	87	80	696
380	4	1165	1383	gi|142570	ATP synthase c subunit [Bacillus firmus]	87	80	219
414	4	900	1073	gi|467386	thiophen and furan oxidation [Bacillus subtilis]	87	77	174
425	2	1003	794	gi|1046166	pilin repressor [Mycoplasma genitalium]	87	69	210
448	1	1255	722	gi|405134	acetate kinase [Bacillus subtilis]	87	75	534
480	1	1	711	gi|142559	ATP synthase alpha subunit [Bacillus megaterium]	87	79	711
481	1	2	352	sp|Q06797|RL1_B	50S RIBOSOMAL PROTEIN L1 (BL1).	87	72	351
677	2	359	955	gi|460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	597
677	3	934	1284	gi|460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	351
876	1	3	452	gi|1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	87	79	450

TABLE 2

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1376	1	426	214	gi|1065555	F46H6.4 gene product (Caenorhabditis elegans)	87	75	213
2206	1	1	374	gi|215098	excisionase (Bacteriophage 154e)	87	72	372
2938	1	3	290	gi|508979	GTP-binding protein (Bacillus subtilis)	87	69	288
3081	2	126	308	gi|467399	IMP dehydrogenase (Bacillus subtilis)	87	72	183
3515	1	3	401	gi|1405454	aconitase (Bacillus subtilis)	87	80	399
4238	1	547	275	gi|601769	HutU protein, urocanase (Bacillus subtilis)	87	73	273
4	8	10427	8736	gi|603769	HutU protein, urocanase (Bacillus subtilis)	86	72	1692
22	6	4190	3738	gi|410515	urease beta subunit (Staphylococcus xylosus)	86	73	453
54	2	2480	1572	gi|289287	UDP-glucose pyrophosphorylase (Bacillus subtilis)	86	70	909
124	3	2336	1713	gi|556887	uracil phosphoribosyltransferase (Bacillus subtilis) pir|S49364|S49364 uracil phosphoribosyltransferase - Bacillus subtilis	86	74	624
148	3	1349	3448	gi|467458	cell division protein (Bacillus subtilis)	86	75	2100
148	4	3638	3859	gi|467460	unknown (Bacillus subtilis)	86	73	222
152	3	1340	2086	gi|1337835	pyruvate decarboxylase E-1 alpha subunit (Bacillus subtilis)	86	75	747
164	18	17347	19467	gi|1184680	polynucleotide phosphorylase (Bacillus subtilis)	86	72	2121
180	2	554	1159	gi|143467	ribosomal protein S4 (Bacillus subtilis)	86	80	606
205	3	2966	2592	gi|142464	ribosomal protein L17 (Bacillus subtilis)	86	77	375
205	26	13364	12990	gi|40107	ribosomal protein L22 (Bacillus stearothermophilus) lr|S10612|S10612 ribosomal protein L22 - Bacillus earothermophilus	86	75	375
246	7	3463	3140	gi|467375	ribosomal protein S6 (Bacillus subtilis)	86	70	324
299	3	1196	1540	gi|39656	apoVG gene product (Bacillus megaterium)	86	70	345
299	7	3884	4345	gi|467440	phosphoribosylpyrophosphate synthetase (Bacillus subtilis) gi|40218 PRPP synthetase (AA 1-317) (Bacillus subtilis)	86	78	462
304	5	2170	2523	gi|666983	putative ATP binding subunit (Bacillus subtilis)	86	65	354
310	2	1487	1678	gi|1177684	chorismate mutase (Staphylococcus xylosus)	86	71	192
337	5	2086	3405	gi|487434	isocitrate dehydrogenase (Bacillus subtilis)	86	78	1320
339	2	1489	1109	gi|1118003	dihydroneopterin aldolase (Staphylococcus haemolyticus)	86	77	381
358	2	2124	3440	gi|1146219	28.2% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)	86	73	1317

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% idenc	length (nt)
404	2	1015	2058	gi|1303817	YqfA [Bacillus subtilis]	86	78	1044
581	2	661	452	gi|40056	phoP gene product [Bacillus subtilis]	86	71	210
642	2	338	1075	gi|1176399	EpiF [Staphylococcus epidermidis]	86	72	738
770	1	622	347	gi|143328	phoP protein (put.); putative [Bacillus subtilis]	86	69	276
865	1	1777	890	gi|1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	86	74	888
868	2	963	1133	gi|1002911	transmembrane protein [Saccharomyces cerevisiae]	86	69	171
904	1	1	162	gi|1303912	YqhW [Bacillus subtilis]	86	72	162
989	1	35	433	gi|1303993	YqkL [Bacillus subtilis]	86	76	399
1212	1	296	150	gi|1414014	ipa-90d gene product [Bacillus subtilis]	86	70	147
1323	1	2	148	gi|40041	pyruvate dehydrogenase (lipoamide) [Bacillus stearothermophilus] ir|S10798|DEBSPF pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus	86	75	147
1085	2	540	310	gi|1354211	PET112-like protein [Bacillus subtilis]	86	86	231
1847	1	1	228	gi|296464	ATPase [Lactococcus lactis]	86	63	228
4487	1	476	240	gi|1022726	unknown [Staphylococcus haemolyticus]	86	73	237
4583	1	372	187	gi|1022725	unknown [Staphylococcus haemolyticus]	86	79	186
25	5	4287	5039	gi|1502421	3-ketoacyl-acyl carrier protein reductase [Bacillus subtilis]	85	64	753
56	21	30627	29395	gi|1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	85	69	1233
68	2	332	1192	gi|467376	unknown [Bacillus subtilis]	85	74	861
73	2	880	1707	gi|142992	glycerol kinase (glpK) (EC 2.7.1.30) [Bacillus subtilis] pir|B45868|B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp|P18157|GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	85	72	828
106	4	1505	3490	gi|143766	[thrSv] (EC 6.1.1.3) [Bacillus subtilis]	85	74	1986
128	2	1153	2202	gi|311924	glyceraldehyde-3-phosphate dehydrogenase (Clostridium pasteurianum) pir|S34254|S34254 glyceraldehyde-3-phosphate dehydrogenase (EC .2.1.12) - Clostridium pasteurianum	85	75	1050
129	4	6466	5252	gi|1064807	ORNITHINE AMINOTRANSFERASE [Bacillus subtilis]	85	73	1215
138	6	3475	5673	gi|1072419	glcB gene product [Staphylococcus carnosus]	85	74	2199
189	1	2	169	gi|467385	unknown [Bacillus subtilis]	85	65	168

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
205	15	8624	8106	gi|1044981	ribosomal protein S5 (Bacillus subtilis)	85	75	519
205	20	10928	10596	pir|A02819|R565	ribosomal protein L24 - Bacillus stearothermophilus	85	72	333
220	6	6490	6101	gi|148980	secA gene product (Bacillus subtilis)	85	66	390
231	4	4877	3159	gi|1002520	MutS (Bacillus subtilis)	85	70	1719
243	9	8013	8783	gi|1414011	ipa-87r gene product (Bacillus subtilis)	85	72	771
249	2	5894	3186	gi|1405454	aconitase (Bacillus subtilis)	85	73	2709
302	1	140	475	gi|40173	homolog of E.coli ribosomal protein L21 (Bacillus subtilis) ir|S18439|S18439 Ribosomal protein L21 - Bacillus subtilis p|P26908|RL21_BACSU 50S RIBOSOMAL PROTEIN L21 (BL20).	85	72	336
333	1	5445	2968	gi|442360	ClpC adenosine triphosphatase (Bacillus subtilis)	85	69	2478
164	6	6082	8196	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	85	68	2115
448	2	1992	1339	gi|405134	acetate kinase (Bacillus subtilis)	85	68	654
747	1	1251	853	gi|1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author (Bacillus subtilis)	85	73	399
886	2	159	467	gi|541768	hemin permease (Yersinia enterocolitica)	85	55	309
1089	1	1208	606	pir|B47154|B471	signal recognition particle 54K chain homolog Ffh - Bacillus subtilis	85	71	601
1163	1	816	409	gi|304155	diaminopimelate decarboxylase (Bacillus methanolicus) sp|P41023|DCDA_BACMT DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20) DAP DECARBOXYLASE).	85	62	408
1924	1	487	251	gi|215098	excisionase (Bacteriophage 154)	85	73	237
2932	1	776	390	gi|1041099	Pyruvate Kinase (Bacillus licheniformis)	85	71	387
3010	1	3	275	gi|42370	pyruvate formate-lyase (AA 1-760) (Escherichia coli) ir|S01788|S01788 formate C-acetyltransferase (EC 2.3.1.54) - cherichia coli	85	74	273
3111	1	595	299	gi|63568	limb deformity protein (Gallus gallus)	85	85	297
3778	1	630	316	gi|391840	beta-subunit of HDT (Pseudomonas fragi)	85	67	315
3835	1	1	387	gi|1204472	type I restriction enzyme ECOR124/3 I M protein (Haemophilus influenzae)	85	56	387
4042	1	3	386	gi|1817R	formate acetyltransferase (Chlamydomonas reinhardtii) ir|S24997|S24997 formate C-acetyltransferase (EC 2.3.1.54) - lamydomonas reinhardtii	85	70	384
4053	1	35	340	gi|1204472	type I restriction enzyme ECOR124/3 I M protein (Haemophilus influenzae)	85	56	306
4108	1	2	181	gi|1072418	glcA gene product (Staphylococcus carnosus)	85	61	180
4300	1	575	330	gi|1151932	fructose enzyme II (Rhodobacter capsulatus)	85	59	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name ( )	% sim	% ident	length (nt)
4392	1	627	355	gi|1022725	unknown (Staphylococcus haemolyticus)	85	74	273
4408	1	2	235	gi|871784	Clp-like ATP-dependent protease binding subunit (Bos taurus)	85	62	234
4430	1	578	291	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	85	68	288
4555	1	2	253	gi|450688	hsdM gene of EcoprrI gene product (Escherichia coli) pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	85	52	252
4611	1	481	242	gi|1256635	dihydroxy-acid dehydratase (Bacillus subtilis)	85	65	240
4	10	10061	10591	gi|46982	tcaB gene product (Staphylococcus epidermidis)	84	68	531
13	2	1348	1172	gi|142450	ahrC protein (Bacillus subtilis)	84	56	177
16	4	1803	4652	gi|1277198	DNA repair protein (Deinococcus radiodurans)	84	67	2850
22	3	1535	1128	gi|511069	UreF (Staphylococcus xylosus)	84	73	408
23	7	5055	5306	gi|603320	Yer082p (Saccharomyces cerevisiae)	84	61	252
53	11	11597	11145	gi|1303948	YqiW (Bacillus subtilis)	84	68	453
53	12	14059	12770	gi|142611	branched chain alpha-keto acid dehydrogenase E2 (Bacillus subtilis) gi|1303944 BfmBB (Bacillus subtilis)	84	71	1290
70	1	1332	982	gi|46647	ORF (repB) (Staphylococcus aureus)	84	68	351
73	4	2512	4311	gi|142993	glycerol-3-phosphate dehydrogenase (glpD) (EC 1.1.99.5) (Bacillus subtilis)	84	74	1800
98	7	4324	6096	gi|467427	methionyl-tRNA synthetase (Bacillus subtilis)	84	66	1773
100	9	9501	8680	gi|1340128	ORF1 (Staphylococcus aureus)	84	78	822
117	3	1934	3208	gi|1237019	Srb (Bacillus subtilis)	84	68	1275
148	6	4720	5670	gi|467462	cysteine synthetase A (Bacillus subtilis)	84	69	951
152	4	2064	2456	gi|141377	pyruvate decarboxylase (E-1) alpha subunit (Bacillus subtilis) pir|B36718|BEDSPA pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) lpha chain - Bacillus subtilis	84	70	393
169	7	3634	3861	gi|1001342	hypothetical protein (Synechocystis sp.)	84	66	228
171	4	2992	2657	gi|517475	D-amino acid transaminase (Staphylococcus haemolyticus)	84	71	336
186	6	6941	6216	gi|467475	unknown (Bacillus subtilis)	84	70	726
205	9	6261	5692	gi|216340	ORF for adenylate kinase (Bacillus subtilis)	84	71	570
224	2	915	1391	gi|288269	beta-fructofuranosidase (Staphylococcus xylosus)	84	70	477

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% idmt	length (nt)
251	1	92	388	gi|1303790	Yqet (Bacillus subtilis)	84	65	297
282	3	1526	2836	gi|143040	glutamate-1-semialdehyde 2,1-aminotransferase (Bacillus subtilis) pir|D42728|D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC .4.3.8) - Bacillus subtilis	84	75	1311
307	5	3138	2959	gi|1070014	protein-dependent (Bacillus subtilis)	84	62	180
320	4	2343	4229	gi|143390	carbamyl phosphate synthetase (Bacillus subtilis)	84	70	1887
372	1	3	296	gi|1022725	unknown (Staphylococcus haemolyticus)	84	70	294
413	2	2201	1341	gi|1256146	YbbQ (Bacillus subtilis)	84	65	861
439	1	3	392	gi|1046173	osmotically inducible protein (Mycoplasma genitalium)	84	53	390
461	3	1162	2270	gi|40211	threonine synthase (thrC) (AA 1-352) (Bacillus subtilis) |r|A25364|A25364 threonine synthase (EC 4.2.99.2) - Bacillus btilis	84	69	909
487	1	3	299	gi|1144531	integrin-like protein alpha Intip (Candida albicans)	84	46	297
491	2	624	905	pir|S08564|R3BS	ribosomal protein S9 - Bacillus stearothermophilus	84	69	282
491	3	836	1033	pir|S08564|R3BS	ribosomal protein S9 - Bacillus stearothermophilus	84	77	198
548	1	3	341	gi|41123?	uracil permease (Bacillus caldolyticus)	84	74	339
728	2	2201	1748	gi|912445	DNA polymerase (Bacillus caldotenax)	84	68	954
769	1	3	257	gi|1510953	cobalamin biosynthesis protein N (Methanococcus jannaschii)	84	38	255
954	1	308	156	gi|1405454	aconitase (Bacillus subtilis)	84	57	151
957	1	3	395	gi|143402	recombination protein (ttg start codon) (Bacillus subtilis) gi|1303923 RecN (Bacillus subtilis)	84	68	393
975	1	3	452	gi|885934	ClpB (Synechococcus sp.)	84	70	450
1385	1	3	257	gi|510140	oligoendopeptidase F (Lactococcus lactis)	84	56	255
2954	1	3	323	gi|603769	HutU protein, urocanase (Bacillus subtilis)	84	73	321
2996	1	650	348	gi|18178	formate acetyltransferase (Chlamydomonas reinhardtii) |r|S24997|S24997 formate C-acetyltransferase (EC 2.3.1.54) - lenydomonas reinhardtii	84	65	303
3766	1	737	375	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	84	72	363
4022	1	2	169	gi|1146206	glutamate dehydrogenase (Bacillus subtilis)	84	54	168
4058	1	620	312	gi|151932	fructose enzyme II (Rhodobacter capsulatus)	84	71	309
4108	2	106	351	gi|1072418	glcA gene product (Staphylococcus carnosus)	84	77	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4183	1	3	308	gi|603769	HutU protein, urocanase [Bacillus subtilis]	84	72	306
4726	1	55	234	gi|146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pir|A29617|A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large chain - Escherichia coli	84	73	180
22	4	2043	1576	gi|393297	urease accessory protein [Bacillus sp.]	83	64	468
53	13	14722	13745	gi|142612	branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus subtilis]	83	68	978
57	16	13357	12872	gi|143132	lactate dehydrogenase (AC 1.1.1.27) [Bacillus caldolyticus] pir|B29704|B29704 L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus aldolyticus	83	66	486
66	3	3119	2274	gi|1301894	YqhM [Bacillus subtilis]	83	63	846
66	5	6118	4643	gi|1212730	YqhX [Bacillus subtilis]	83	68	1476
70	3	1864	1523	gi|44095	replication initiator protein [Listeria monocytogenes]	83	73	342
90	1	377	1429	gi|155571	alcohol dehydrogenase I (adhA) (EC 1.1.1.1) [Zymomonas mobilis] pir|A35260|A35260 alcohol dehydrogenase (EC 1.1.1.1) I - Zymomonas mobilis	83	70	1053
95	2	708	2162	gi|506381	phospho-beta-glucosidase [Bacillus subtilis]	83	70	1455
137	1	68	694	gi|467391	initiation protein of replicaton [Bacillus subtilis]	83	77	627
140	4	3209	2742	gi|634107	kdpB [Escherichia coli]	83	65	468
142	3	3468	2989	gi|1212775	lumazine synthase (b-subunit) [Bacillus amyloliquefaciens]	83	69	480
161	12	5749	6696	gi|903307	ORF75 [Bacillus subtilis]	83	64	948
164	9	9880	11070	gi|149316	ORF2 gene product [Bacillus subtilis]	83	66	1191
164	14	14148	14546	gi|580902	ORF6 gene product [Bacillus subtilis]	83	60	399
170	2	3144	2467	gi|520844	orf4 [Bacillus subtilis]	83	64	678
186	2	2029	1370	gi|289284	cysteinyl-tRNA synthetase [Bacillus subtilis]	83	72	660
205	14	7822	7607	gi|216337	ORF for L30 ribosomal protein [Bacillus subtilis]	83	74	216
237	6	3683	4540	gi|1510488	imidazoleglycerol-phosphate synthase (cyclase) [Methanococcus jannaschii]	83	60	858
301	1	985	638	gi|1467419	unknown [Bacillus subtilis]	83	65	348
302	4	1421	2743'	gi|508979	GTP-binding protein [Bacillus subtilis]	83	68	1323
321	4	3933	3571	gi|139844	fumarase (citG) (aa 1-462) [Bacillus subtilis]	83	68	363
367	1	2	352	gi|1039479	ORFU [Lactococcus lactis]	83	54	351

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
147	1	3	662	gi|806281	DNA polymerase I [Bacillus stearothermophilus]	83	70	660
527	2	916	1566	gi|396259	protease [Staphylococcus epidermidis]	83	67	651
533	1	355	179	gi|142055	alanine dehydrogenase (EC 1.4.1.1) [Bacillus stearothermophilus] pir|B34261|B34261 alanine dehydrogenase (EC 1.4.1.1) - Bacillus tearothermophilus	83	66	177
536	4	1617	1438	gi|143366	adenylosuccinate.lyase (PUR-B) [Bacillus subtilis] ;ir|C29326|W2BSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis	83	67	180
652	1	2	859	gi|520753	DNA topoisomerase I [Bacillus subtilis]	83	72	858
774	2	200	361	gi|1522665	M. jannaschii predicted coding region MJECL28 [Methanococcus jannaschii]	83	58	162
897	1	120	296	gi|1064807	ORNITHINE AMINOTRANSFERASE [Bacillus subtilis]	83	76	177
1213	1	3	491	gi|289288	lexA [Bacillus subtilis]	83	67	489
2529	1	296	150	gi|143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pir|JT0481|YWBS tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus ubtilis	83	69	147
2973	1	649	326	gi|1109687	ProZ [Bacillus subtilis]	83	58	324
1009	1	728	366	gi|882532	ORF_o294 [Escherichia coli]	83	65	363
3035	2	45	305	gi|950062	hypothetical yeast protein 1 (Mycoplasma capricolum) pir|S48578|S48578 hypothetical protein - Mycoplasma capricolum SGC3) (fragment)	83	59	261
3906	1	67	309	gi|1353197	thioredoxin reductase [Eubacterium acidaminophilum]	83	61	243
445A	1	540	271	gi|1397526	clumping factor [Staphylococcus aureus]	83	78	270
4570	1	444	223	gi|1022726	unknown [Staphylococcus haemolyticus]	83	74	222
4654	1	97	261	gi|1072419	glcB gene product [Staphylococcus carnosus]	83	79	165
16	2	295	1191	gi|153854	uvs402 protein [Streptococcus pneumoniae]	82	67	897
16	3	1193	1798	gi|153854	uvs402 protein [Streptococcus pneumoniae]	82	70	606
3A	12	9644	8724	gi|1204400	N-acetylneuraminate lyase [Haemophilus influenzae]	82	58	921
42	4	988	2019	gi|841192	catalase [Bacteroides fragilis]	82	70	1032
51	6	2590	3489	gi|143607	sporulation protein [Bacillus subtilis]	82	69	900
56	11	12270	13925	gi|39431	oligo-1,6-glucosidase [Bacillus cereus]	82	60	1656
56	15	17673	18014	gi|467410	unknown [Bacillus subtilis]	82	66	342
61	2	881	3313	gi|143148	transfer RNA-Leu synthetase [Bacillus subtilis]	82	70	2433

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
A2	7	9162	11318	gi|48240	elongation factor G (AA 1-691) (Thermus aquaticus thermophilus) lr|S15928|EFTWG translation elongation factor G - Thermus aquaticus p|P13551|EFG_THETH ELONGATION FACTOR G (EF-G).	82	64	2157
85	2	5470	3260	gi|143369	phosphoribosylformyl glycinamidine synthetase II (PUR-Q) [Bacillus subtilis]	82	66	2211
102	6	3662	5380	gi|1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	82	65	1719
117	4	3242	3493	pir|A47154|A471	orf1 5' of Frh - Bacillus subtilis	82	55	252
128	6	4377	5933	gi|460258	phosphoglycerate mutase [Bacillus subtilis]	82	66	1557
129	2	1229	2182	gi|40377	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] plr|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acilius subtilis	82	62	954
170	1	2	1441	gi|1377831	unknown [Bacillus subtilis]	82	67	1440
177	1	3	1094	gi|467386	thiophen and furan oxidation [Bacillus subtilis]	82	65	1092
184	4	3572	4039	gi|153566	ORF (19K protein) (Enterococcus faecalis)	82	59	468
189	8	4455	4225	gi|1001878	CapL protein [Listeria monocytogenes]	82	71	231
206	19	21366	20707	gi|473916	lipopeptide antibiotics iturin A [Bacillus subtilis] sp|P39144|LP14_BACSU LIPOPEPTIDE ANTIBIOTICS ITURIN A AND SURFACTIN IOSYNTHESIS PROTEIN.	82	50	660
221	2	805	1722	gi|517205	67 kDa Hyosin-crossreactive streptococcal antigen [Streptococcus yogenes]	82	63	918
223	4	3866	3651	gi|439619	[Salmonella typhimurium IS200 insertion sequence from SARA17, artial.], gene product [Salmonella typhimurium]	82	69	216
260	3	5207	4296	gi|1161381	IcaB [Staphylococcus epidermidis]	82	61	912
315	3	4864	2855	gi|143397	quinol oxidase [Bacillus subtilis]	82	67	2010
321	10	8520	7945	gi|1429R1	ORF5; This ORF includes a region (aa23-103) containing a potential non-sulphur centre homologous to a region of Rhodospirillum rubrum nd Chromatium vinosum; putative [Bacillus stearothermophilus] pir|PQ0299|PQ0299 hypothetical protein 5 (glnA 3' region) -	82	62	576
331	3	1055	1342	gi|435574	ribosomal protein L1 [Bacillus subtilis]	82	71	288
370	2	262	618	gi|1303793	YqeL [Bacillus subtilis]	82	59	357
404	4	3053	4024	gi|1303821	YqfE [Bacillus subtilis]	82	68	972
405	4	4440	3073	gi|1303913	YqhX [Bacillus subtilis]	82	67	1368
436	1	4096	2864	gi|149521	tryptophan synthase beta subunit [Lactococcus lactis] pir|S35129|S35129 tryptophan synthase (EC 4.2.1.20) beta chain - actococcus lactis subsp. lactis	82	67	1233

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name		% sim	% ident	length (nt)
441	4	3394	2573	gi|142952	glyceraldehyde-3-phosphate dehydrogenase (Bacillus stearothermophilus)		82	67	822
444	12	10415	11227	gi|1204354	spore germination and vegetative growth protein (Haemophilus influenzae)		82	67	813
446	1	3	191	gi|143387	aspartate transcarbamylase (Bacillus subtilis)		82	66	189
462	3	1007	1210	gi|142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] pir|A37192|A37192 uvrB protein - Bacillus subtilis sp|P14951|UVRC_BACSU EXCINUCLEASE ABC SUBUNIT C.		82	64	204
537	1	1560	784	gi|853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]		82	61	777
680	2	407	700	gi|426472	secK gene product (Staphylococcus carnosus)		82	69	294
724	2	565	386	gi|143377	phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-HIJ) Bacillus subtilis)		82	68	180
763	1	422	213	gi|467458	cell division protein [Bacillus subtilis]		82	35	210
818	1	564	283	gi|1064787	function unknown [Bacillus subtilis]		82	69	282
858	1	175	1176	gi|143043	uroporphyrinogen decarboxylase [Bacillus subtilis] pir|B47045|B47045 uroporphyrinogen decarboxylase (EC 4.1.1.37) - acillus subtilis		82	71	1002
895	1	3	599	gi|1027507	ATP binding protein (Borrelia burgdorferi)		82	72	597
939	1	10	399	gi|143795	transfer RNA-Tyr synthetase (Bacillus subtilis)		82	60	390
961	1	1	306	gi|577647	gamma-hemolysin (Staphylococcus aureus)		82	69	306
1192	1	107	155	gi|146974	NAD-dependent NAD synthetase [Escherichia coli]		82	71	153
1317	1	49	375	gi|407908	EIIacr (Staphylococcus xylosus)		82	72	327
1341	1	1	150	gi|39962	ribosomal protein L35 (AA 1-66) [Bacillus stearothermophilus] ir|S05347|R58S35 ribosomal protein L35 - Bacillus earothermophilus		82	68	150
2990	2	567	349	gi|534855	ATPase subunit epsilon [Bacillus stearothermophilus] sp|P42009|ATPE_BACST ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).		82	47	219
3024	1	45	224	gi|467402	unknown [Bacillus subtilis]		82	64	180
3045	1	276	139	gi|467335	ribosomal protein L9 [Bacillus subtilis]		82	60	138
3045	2	558	400	gi|467335	ribosomal protein L9 [Bacillus subtilis]		82	82	159
3091	1	474	238	gi|499335	secA protein (Staphylococcus carnosus)		82	78	237
3107	1	416	210	gi|546918	orf(Y 3' of comK [Bacillus subtilis, E26, Peptide Partial, 140 aa] pir|S43612|S43612 hypothetical protein Y - Bacillus subtilis sp|P40398|YHXD_BACSU HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFY) FRAGMENT).		82	64	207

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4332	1	2	319	gi|42086	nitrate reductase alpha subunit [Escherichia coli] p|P09152|NARG_ECOLI RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4). (SUB 2-1247)	82	75	318
23	3	3275	2574	gi|1199573	spsB [Sphingomonas sp.]	81	64	702
42	1	638	321	gi|466778	lysine specific permease (Escherichia coli)	81	59	318
48	5	4051	4350	gi|1045937	M. genitalium predicted coding region MG246 [Mycoplasma genitalium]	81	62	300
51	4	1578	2579	pir|S16649|S166	dciAC protein - Bacillus subtilis	81	55	1002
53	2	364	1494	gi|1303961	YqjU [Bacillus subtilis]	81	67	1131
53	8	9419	7971	gi|146930	6-phosphogluconate dehydrogenase [Escherichia coli]	81	66	1449
54	9	10757	10119	gi|143016	permease [Bacillus subtilis]	81	65	639
54	10	13360	11786	gi|143015	gluconate kinase [Bacillus subtilis]	81	64	1575
57	17	13983	13366	pir|A25805|A258	L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	81	74	618
81	2	2708	2217	gi|1222302	NifU-related protein [Haemophilus influenzae]	81	54	492
86	1	745	374	gi|414017	ipa-93d gene product [Bacillus subtilis]	81	70	372
103	6	6438	4861	gi|971342	nitrate reductase beta subunit [Bacillus subtilis] sp|P42176|NARH_BACSU NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4).	81	64	1578
120	15	10845	12338	gi|1524392	GbsA [Bacillus subtilis]	81	67	1494
128	5	3676	4413	gi|1143319	triose phosphate isomerase (Bacillus megaterium)	81	64	738
131	9	10308	9280	gi|1299163	alanine dehydrogenase [Bacillus subtilis]	81	68	1029
143	6	6088	5471	gi|1439619	[Salmonella typhimurium IS200 insertion sequence from SARA17, artia1.1, gene product (Salmonella typhimurium)]	81	61	618
169	1	43	825	gi|897795	30S ribosomal protein (Pediococcus acidilactici) sp|P49668|RS2_PEDAC 30S RIBOSOMAL PROTEIN S2.	81	65	783
230	1	450	226	gi|1125826	short region of weak similarity to tyrosine-protein kinase receptors in a fibronectin type III-like domain (Caenorhabditis elegans)	81	54	225
233	5	2000	2677	gi|467404	unknown [Bacillus subtilis]	81	63	678
241	2	3081	2149	gi|16510	succinate--CoA ligase (GDP-forming) (Arabidopsis thaliana) tr|S30579|S30579 succinate--CoA ligase (GDP-forming) (EC 6.2.1.4) pha chain - Arabidopsis thaliana (fragment)	81	69	933
256	1	1	981	pir|S09411|S094	spoIIIE protein - Bacillus subtilis	81	65	981
259	3	3752	2691	sp|P28367|RF2_B	PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT).	81	65	1062

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name (; ...)	% sim	% ident	Length (nt)
275	2	1728	3581	gi|726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis]	81	68	1854
285	1	1466	735	gi|1204844	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	81	63	732
296	1	99	1406	gi|467328	adenylosuccinate synthetase [Bacillus subtilis]	81	67	1308
302	9	5590	5689	gi|147485	queA [Escherichia coli]	81	64	100
317	2	1137	1376	gi|154961	resolvase [Transposon Tn917]	81	5*	240
343	2	1034	1342	gi|405955	yeeD [Escherichia coli]	81	60	109
360	2	1404	2471	gi|1204570	aspartyl-tRNA synthetase [Haemophilus influenzae]	81	67	1068
364	5	6351	5706	gi|1204652	methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]	81	63	546
372	2	1707	1135	gi|467416	unknown [Bacillus subtilis]	81	65	573
392	1	43	603	pir|S09411|S094	spoIIE protein - Bacillus subtilis	81	65	561
404	9	5252	6154	gi|1606745	Bex [Bacillus subtilis]	81	65	903
426	2	1727	1119	gi|39453	Manganese superoxide dismutase [Bacillus caldotenax] (r|S22053|S22053) superoxide dismutase (EC 1.15.1.1) (Mn) - Bacillus ldotenax	81	66	609
480	?	4653	5889	pir|C370MJ|C370	hypothetical protein II (ompII 3' region) - Salmonella typhimurium (fragment)	81	57	217
625	3	1105	2070	gi|1263360	protein kinase PknB [Mycobacterium leprae]	81	56	966
754	2	504	1064	gi|1303902	YqhU [Bacillus subtilis]	81	71	561
842	1	86	430	gi|1405446	transketolase [Bacillus subtilis]	81	68	345
953	1	798	400	gi|1205429	dipeptide transport ATP-binding protein [Haemophilus influenzae]	81	57	399
961	2	252	401	gi|467686	synergohymenotropic toxin [Staphylococcus intermedius] pir|S44944|S44944 synergohymenotropic toxin - Staphylococcus intermedius	81	72	150
1035	1	1	189	gi|1046138	M. genitalium predicted coding region MG423 [Mycoplasma genitalium]	81	43	189
1280	1	670	449	gi|559164	helicase [Autographa californica nuclear polyhedrosis virus] sp|P24307|V143_NPVAC HELICASE.	81	43	222
3371	1	68	241	gi|1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	81	62	174
3715	1	475	239	gbf|537137	ORF_f388 [Escherichia coli]	81	58	237
3908	1	2	325	gi|439619	[Salmonella typhimurium IS200 insertion sequence from SARA17, artial.]. gene product [Salmonella typhimurium]	81	68	324
3940	1	3	401	gbf|296464	ATPase [Lactococcus lactis]	81	69	399

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3954	1	1	318	gi|1224069	amidase [Moraxella catarrhalis]	81	68	318
4049	1	337	170	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis	81	68	168
4209	1	1	324	gi|403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	81	58	324
4371	1	627	322	gi|216677	indolepyruvate decarboxylase [Enterobacter cloacae] pir|S16013|S16013 indolepyruvate decarboxylase (EC 4.1.1.-) - nterobacter cloacae	81	72	306
4387	1	19	228	gi|460689	TVG [Thermoactinomyces vulgaris]	81	59	210
4391	1	581	306	gi|1524193	unknown [Mycobacterium tuberculosis]	81	67	276
4425	1	3	341	gi|143015	gluconate kinase [Bacillus subtilis]	81	66	339
9	1	1593	847	gi|1064786	function unknown [Bacillus subtilis]	80	62	747
17	1	544	311	gi|559164	helicase [Autographa californica nuclear polyhedrosis virus] sp|P24307|V143_NPVAC HELICASE.	80	40	234
45	2	1159	2448	gi|1109684	ProV [Bacillus subtilis]	80	63	1290
45	5	4032	4733	gi|1109687	ProZ [Bacillus subtilis]	80	55	702
54	8	10266	9502	gi|563952	gluconate permease [Bacillus licheniformis]	80	62	765
62	12	8852	7545	gi|854655	Na/H antiporter system [Bacillus alcalophilus]	80	62	1308
62	14	8087	8683	gi|559713	ORF [Homo sapiens]	80	68	597
67	16	13781	14122	gi|305002	ORF_4356 [Escherichia coli]	80	65	342
70	13	11495	10296	gi|1303995	YqkN [Bacillus subtilis]	80	64	1200
98	9	6336	7130	gi|467428	unknown [Bacillus subtilis]	80	68	795
98	10	7294	7833	gi|467430	unknown [Bacillus subtilis]	80	64	540
98	11	7820	8737	gi|467431	high level kasagamycin resistance [Bacillus subtilis]	80	61	918
109	16	14154	14813	gi|580875	ipa-57d gene product [Bacillus subtilis]	80	63	660
112	15	14294	16636	gi|1072361	pyruvate-formate-lyase [Clostridium pasteurianum]	80	65	2343
139	1	1448	726	gi|506699	CapC [Staphylococcus aureus]	80	58	723
139	2	2179	1448	gi|505698	CapB [Staphylococcus aureus]	80	59	732
174	4	3271	2870	gi|1146242	aspartate 1-decarboxylase [Bacillus subtilis]	80	61	402

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
177	3	2102	2842	gi|467385	unknown [Bacillus subtilis]	80	70	741
184	6	6124	5912	gi|161953	85-kDa surface antigen [Trypanosoma cruzi]	80	46	213
186	4	5368	3875	gi|289282	glutamyl-tRNA synthetase [Bacillus subtilis]	80	65	1494
205	10	15796	15140	gi|40103	ribosomal protein L4 [Bacillus stearothermophilus]	80	66	657
207	1	140	1315	gi|460259	enolase [Bacillus subtilis]	80	67	1176
211	3	1078	1590	gi|410131	ORFX7 [Bacillus subtilis]	80	61	513
235	2	1962	2255	gi|143797	valyl-tRNA synthetase [Bacillus stearothermophilus] sp|P19931|SYV_BACST VALVL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE) (VALRS).	80	55	294
239	1	1	1263	gi|143000	proton glutamate symport protein [Bacillus stearothermophilus] pir|S26247|S26247 glutamate/aspartate transport protein - Bacillus stearothermophilus	80	59	1263
272	5	2724	2461	gi|709993	hypothetical protein [Bacillus subtilis]	80	54	264
301	3	1446	1111	gi|467418	unknown [Bacillus subtilis]	80	58	336
310	4	5697	4501	gi|1177686	acuC gene product [Staphylococcus xylosus]	80	67	1197
310	6	5258	7006	gi|348053	acetyl-CoA synthetase [Bacillus subtilis]	80	67	1749
310	7	7410	9113	gi|1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	80	67	1704
325	3	1114	1389	gi|310225	outer capsid protein [Rotavirus sp.]	80	40	276
337	1	1268	636	gi|537049	ORF_o470 [Escherichia coli]	80	55	633
374	2	929	1228	gi|1405448	YneP [Bacillus subtilis]	80	70	300
375	5	3062	3331	gi|467448	unknown [Bacillus subtilis]	80	68	270
388	1	267	587	gi|1064791	function unknown [Bacillus subtilis]	80	65	321
394	1	9	659	gi|304976	matches PS00017: ATP_GTP_A and PS00301: EFACTOR_GTP; similar to longation [factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli]	80	65	651
456	1	625	1263	gi|1146183	putative [Bacillus subtilis]	80	65	639
475	1	1	654	gi|286269	beta-fructofuranosidase [Staphylococcus xylosus]	80	66	654
544	2	1449	2240	gi|529754	speC [Streptococcus pyogenes]	80	50	792
622	4	1623	1871	gi|1483545	unknown [Mycobacterium tuberculosis]	80	65	249
719	1	1	1257	gi|1064791	function unknown [Bacillus subtilis]	80	68	1257
739	1	107	838	gi|1666983	putative ATP binding subunit [Bacillus subtilis]	80	61	732

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
745	2	581	414	gi|1511600	coenzyme PQQ synthesis protein III [Methanococcus jannaschii]	80	61	168
822	1	17	679	gi|410141	ORFX17 [Bacillus subtilis]	80	68	663
A27	2	991	836	pi|1205301	leukotoxin secretion ATP-binding protein [Haemophilus influenzae]	80	54	156
1044	1	3	149	gi|60632	vp2 [Marburg virus]	80	55	147
1220	2	571	413	pir|A61072|EPSG	gallidermin precursor - Staphylococcus gallinarum	80	78	159
2519	1	75	275	gi|147556	dpj [Escherichia coli]	80	45	201
2947	1	503	279	gi|184680	polynucleotide phosphorylase [Bacillus subtilis]	80	62	225
3120	1	2	226	gi|517205	67 kDa Hyosin-crossreactive streptococcal antigen [Streptococcus yogenes]	80	65	225
3193	1	294	148	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	80	59	147
3560	2	285	434	gi|217130	photosystem I core protein B [Synechococcus vulcanus]	80	70	150
3655	1	47	346	gi|415855	deoxyribose aldolase [Mycoplasma hominis]	80	56	100
3658	2	324	584	gi|551531	2-nitropropane dioxygenase [Williopsis saturnus]	80	54	261
3769	1	798	400	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	399
3781	1	692	348	gi|166412	NADH-glutamate synthase [Medicago sativa]	80	62	345
3988	1	48	287	gi|1204696	fructose-permease IIBC component [Haemophilus influenzae]	80	69	240
4030	1	571	287	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	80	60	285
4092	1	547	275	gi|1270207	orf6 [Lactobacillus sake]	80	69	273
4103	1	680	342	gi|39956	IIGlc [Bacillus subtilis]	80	65	339
4231	1	692	348	gi|289287	UDP-glucose pyrophosphorylase [Bacillus subtilis]	80	65	345
4265	1	595	299	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis	80	63	297
4504	1	498	250	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	249
2	6	5998	6798	n|535351	CodY [Bacillus subtilis]	79	63	801
4	7	8295	705f	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis	79	64	1245
25	6	5273	5515	pir|A36728|A367	acyl carrier protein - Rhizobium meliloti	79	65	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
59	2	1173	1424	gi|147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	79	75	252
60	1	1	204	gi|666115	orf1 upstream of glucose kinase [Staphylococcus xylosus] pir|S52351|S52351 hypothetical protein 1 - Staphylococcus xylosus	79	60	204
81	1	3002	1590	gi|466882	ppsI; B1496_C2_169 [Mycobacterium leprae]	79	64	1413
85	7	7023	6505	gi|143364	phosphoribosyl aminoimidazole carboxylase I (PUR-E) [Bacillus subtilis]	79	60	519
89	6	5660	4554	gi|144906	product homologous to B.coli thioredoxin reductase: J.Biol.Chem. 1988) 263:9015-9019, and to F52a protein of alkyl hydroperoxide reductase from S.typhimurium: J.Biol.Chem. (1990) 265:10535-10540; pan reading frame A [Clostridium pasteurianum]	79	35	1107
102	11	7489	8571	gi|143093	ketol-acid reductoisomerase [Bacillus subtilis] sp|P37253|ILVC_BACSU KETOL-ACID REDUCTOISOMERASE (EC 1.1.1.86) ACETOHYDROXY-ACID EDUCTOISOMERASE) (ALPHA-KETO-BETA-HYDROXYLACIL EDUCTOISOMERASE)	79	64	1083
102	14	11190	12563	gi|149428	putative [Lactococcus lactis]	79	65	1374
127	9	7792	9372	gi|458688	PrfC/RF3 [Dichelobacter nodosus]	79	68	1581
139	3	2540	1983	gi|506697	CapA [Staphylococcus aureus]	79	55	558
144	2	1644	1156	gi|1198296	peptide methionine sulfoxide reductase [Streptococcus pneumoniae]	79	47	489
148	2	529	1098	gi|467457	hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi|467457 hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis]	79	59	570
150	1	965	591	gi|755602	unknown [Bacillus subtilis]	79	61	375
176	1	1039	587	gi|297874	fructose-bisphosphate aldolase [Staphylococcus carnosus] pir|A49943|A49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - taphylococcus carnosus (strain TM300)	79	65	453
186	7	7584	6874	gi|1314298	ORF5; putative Sms protein; similar to Sms proteins from Haemophilus influenzae and Escherichia coli [Listeria monocytogenes]	79	64	711
205	16	8887	8498	gi|1044980	ribosomal protein L18 [Bacillus subtilis]	79	70	390
211	1	1	519	gi|1303994	YqkM [Bacillus subtilis]	79	62	519
223	2	4183	2801	gi|488430	alcohol dehydrogenase 2 [Entamoeba histolytica]	79	60	1383
243	8	8915	7896	gi|580883	ipa-88d gene product [Bacillus subtilis]	79	60	1020
279	4	3721	4329'	gi|413930	ipa-6d gene product [Bacillus subtilis]	79	59	609
300	1	11	1393	gi|403372	glycerol 3-phosphate permease [Bacillus subtilis]	79	62	1383
307	3	2930	1935	gi|950062	hypothetical yeast protein 1 [Mycoplasma capricolum] pir|S48578|S48578 hypothetical protein - Mycoplasma capricolum SGC3) (fragment)	79	60	996

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
352	6	10106	8886	gi|216854	P47K [Pseudomonas chlororaphis]	79	59	1221
412	1	1153	578	gi|143177	putative [Bacillus subtilis]	79	51	576
481	3	621	1124	gi|786163	Ribosomal Protein L10 [Bacillus subtilis]	79	66	504
516	1	702	352	gi|805090	NisF [Lactococcus lactis]	79	48	351
525	2	2457	1426	gi|143371	phosphoribosyl aminoimidazole synthetase (PUR-M) [Bacillus subtilis] pir|H29326|A2BSCL phosphoribosylformylglycinamidine cyclo-ligase EC 6.3.3.1] - Bacillus subtilis	79	61	1032
53A	4	3448	2825	gi|1370207	orf6 [Lactobacillus sake]	79	67	624
570	1	2	421	gi|476160	arginine permease substrate-binding subunit [Listeria monocytogenes]	79	61	420
645	8	2663	3241	gi|153898	transport protein [Salmonella typhimurium]	79	62	579
6A3	1	75	374	gi|1064795	function unknown [Bacillus subtilis]	79	62	300
R16	3	4700	3987	gi|1407784	orf-1; novel antigen [Staphylococcus aureus]	79	62	714
2929	1	3	401	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	79	61	399
2937	1	357	202	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	79	58	156
2940	1	768	385	gi|149429	putative [Lactococcus lactis]	79	72	384
2946	1	570	286	gi|143267	2-oxoglutarate dehydrogenase (odhA; EC 1.2.4.2) [Bacillus subtilis]	79	61	285
2999	1	3	212	gi|710020	nitrite reductase (nirB) [Bacillus subtilis]	79	59	210
3022	1	514	332	gi|450686	3-phosphoglycerate kinase [Thermotoga maritima]	79	61	183
3064	1	3	314	gi|1204436	pyruvate formate-lyase [Haemophilus influenzae]	79	60	312
3083	1	2	220	gi|1149662	hypD gene product [Clostridium perfringens]	79	56	219
3126	1	701	411	g:|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	79	55	291
3181	1	607	326	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	79	59	282
3345	1	3	476	gi|871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	79	63	474
3718	1	536	270	pir|C36889|C368	leuB protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	79	71	267
3724	2	159	401	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	79	64	243
3836	1	608	312	pi|1524193	unknown [Mycobacterium tuberculosis]	79	65	297
3941	1	2	334	gi|415855	deoxyribose aldolase [Mycoplasma hominis]	79	54	333
4113	1	3	341	gi|143015	gluconate kinase [Bacillus subtilis]	79	63	319

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4501	1	406	209	gi|1022726	unknown (Staphylococcus haemolyticus)	79	66	198
4612	1	2	238	gi|460689	TVG (Thermoactinomyces vulgaris)	79	58	237
2	1	2	1213	gi|520753	DNA topoisomerase I (Bacillus subtilis)	78	64	1212
8	2	2266	1220	gi|216151	DNA polymerase (gene L; ttg start codon) [Bacteriophage SPO2] gi|579197 SPO2 DNA polymerase (aa 1-648) [Bacteriophage SPO2] pir|A21498|DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	78	72	1047
9	2	1340	1089	gi|1064787	function unknown (Bacillus subtilis)	78	57	252
32	8	6803	7702	gi|146974	NH3-dependent NAD synthetase (Escherichia coli)	78	63	900
36	4	2941	3138	gi|290503	glutamate permease (Escherichia coli)	78	53	198
53	15	17684	16221	gi|1303941	YqiV (Bacillus subtilis)	78	58	1464
57	14	10520	12067	gi|1072418	glcA gene product (Staphylococcus carnosus)	78	65	1548
66	7	6798	5812	gi|1212729	YqhJ (Bacillus subtilis)	78	67	987
67	4	4029	4376	gi|466612	nlkA (Escherichia coli)	78	71	348
91	9	10058	10942	gi|467380	stage O sporultion (Bacillus subtilis)	78	50	885
102	12	8574	10130	gi|149426	putative (Lactococcus lactis)	78	61	1557
112	6	3540	4463	gi|854234	cymG gene product (Klebsiella oxytoca)	78	56	924
124	2	1888	1061	gi|405622	unknown (Bacillus subtilis)	78	60	828
130	3	1805	2260	gi|1256636	putative (Bacillus subtilis)	78	71	456
133	1	751	377	gi|168060	lanB (Emericella nidulans)	78	59	375
166	4	7125	6163	gi|451216	Mannosephosphate isomerase (Streptococcus mutans)	78	63	963
186	1	1586	795	gi|289284	cysteinyl-tRNA synthetase (Bacillus subtilis)	78	63	792
195	4	2749	2315	gi|1353874	unknown (Rhodobacter capsulatus)	78	58	435
199	3	4279	3623	gi|143525	succinate dehydrogenase cytochrome b-558 subunit (Bacillus subtilis) pir|A29843|DEBSSC succinate dehydrogenase (EC 1.3.99.1) cytochrome 558 - Bacillus subtilis	78	57	657
199	4	7209	5557	gi|142521	deoxyribodipyrimidine photolyase (Bacillus subtilis) pir|A37192|A37192 uvrB protein - Bacillus subtilis sp|P14951|UVRC_BACSU EXCINUCLEASE ABC SUBUNIT C.	78	62	1653
223	3	1831	1523	gi|139596	[Escherichia coli IS200 insertion sequence from ECOR63, partial.], ene product [Escherichia coli]	78	47	309

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
299	4	1865	2149	gi|467439	temperature sensitive cell division [Bacillus subtilis]	78	62	285
321	9	7734	7315	gi|142979	ORF3 is homologous to an ORF downstream of the spoT gene of E.coli; RP3 (Bacillus stearothermophilus)	78	55	420
352	4	3714	3944	gi|349050	lectin 1 (Pneumocystis carinii)	78	42	231
352	5	7592	6093	gi|903587	NADH dehydrogenase subunit 5 [Bacillus subtilis] sp|P39755|NDHF_BACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5).	78	58	1500
376	1	2	583	gi|551693	dethiobiotin synthase [Bacillus sphaericus]	78	34	582
424	2	1595	1768	gi|1524117	alpha-acetolactate decarboxylase (Lactococcus lactis)	78	68	174
450	1	1914	988	gi|1030068	NAD(P)H oxidoreductase, isoflavone reductase homologue [Solanum tuberosum]	78	61	927
558	1	762	562	gi|1511588	bifunctional protein [Methanococcus jannaschii]	78	60	201
670	3	1152	1589	gi|1122759	unknown [Bacillus subtilis]	78	64	438
714	1	64	732	gi|143460	37 kd minor sigma factor (rpoF, sigB; ttg start codom) [Bacillus subtilis]	78	57	669
814	1	3	368	gi|1377833	unknown [Bacillus subtilis]	78	59	366
981	1	1381	692	gi|143802	GerC2 [Bacillus subtilis]	78	64	690
995	2	978	727	gi|296947	uridine kinase [Escherichia coli]	78	64	252
1045	1	3	401	gi|1407784	orf-1; novel antigen [Staphylococcus aureus]	78	61	399
1163	2	368	186	gi|141117	diaminopimelate decarboxylase [Bacillus subtilis]	78	54	183
2191	1	794	399	gi|215098	excisionase [Bacteriophage 154a]	78	65	396
2933	1	2	181	gi|1204436	pyruvate formate-lyase [Haemophilus influenzae]	78	73	180
3041	2	129	317	gi|624632	GltL [Escherichia coli]	78	53	189
3581	1	105	401	gi|763186	3-ketoacyl-coA thiolase [Saccharomyces cerevisiae]	78	55	297
3709	1	3	230	gi|460689	TVG [Thermoactinomyces vulgaris]	78	58	228
3974	1	528	265	g-|558639	unknown [Bacillus subtilis]	78	65	264
3980	1	3	401	gi|39956	IIOlc [Bacillus subtilis]	78	62	399
4056	1	647	354	gi|1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	78	55	294
4114	1	630	316	pir|S09372|S093	hypothetical protein - Trypanosoma brucei	78	62	315
4185	1	3	179	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	78	58	177

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4235	1	655	329	gi|558839	unknown [Bacillus subtilis]	78	60	327
4352	1	541	302	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis	78	63	240
4368	1	612	307	gi|1353678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	78	59	306
4461	1	428	216	gi|1276841	glutamate synthase (GOGAT) [Porphyra purpurea]	78	36	213
4530	1	474	238	gi|39956	IIGlc [Bacillus subtilis]	78	65	237
3	2	2969	2073	gi|1109684	ProV [Bacillus subtilis]	77	56	897
12	2	2426	1965	gi|467335	ribosomal protein L9 [Bacillus subtilis]	77	59	462
27	1	2	388	gi|1212728	YqhI [Bacillus subtilis]	77	63	387
39	2	590	1252	gi|40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus btilis]	77	60	663
42	6	2704	2931	gi|606241	30S ribosomal subunit protein S14 [Escherichia coli] sp|P02370|RS14_ECOLI 30S RIBOSOMAL PROTEIN S14 (SUB 2-101)	77	65	228
46	1R	15459	16622	gi|297798	mitochondrial [formate dehydrogenase precursor [Solanum tuberosum] pir|JQ2272|JQ2272 formate dehydrogenase (EC 1.2.1.2) precursor, mitochondrial - potato	77	55	1164
100	4	4562	4002	gi|1340128	ORF1 [Staphylococcus aureus]	77	54	561
102	8	5378	5713	gi|1311482	acetolactate synthase [Thermus aquaticus]	77	57	336
109	7	4742	5383	gi|710637	Unknown [Bacillus subtilis]	77	56	642
117	1	2	1228	gi|1237015	ORF4 [Bacillus subtilis]	77	53	1227
124	10	8323	7688	gi|405819	thymidine kinase [Bacillus subtilis]	77	63	636
147	3	1146	985	gi|849027	hypothetical 15.9-kDa protein [Bacillus subtilis]	77	37	162
152	10	7354	7953	gi|1205583	spermidine/putrescine transport ATP-binding protein [Haemophilus influenzae]	77	55	600
169	2	1004	1282	gi|473825	elongation factor EF-Ts' [Escherichia coli]	77	58	279
184	2	380	1147	gi|216314	esterase [Bacillus stearothermophilus]	77	60	768
189	7	1296	3868	gi|853809	ORF3 [Clostridium perfringens]	77	48	573
193	1	132	290	gi|1303788	YqeH [Bacillus subtilis]	77	54	159
195	8	8740	8414	gi|1499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	77	44	327
205	8	5428	5204	gi|216340	ORF for adenylate kinase [Bacillus subtilis]	77	61	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
205	29	14795	14502	gi|786155	Ribosomal Protein L23 (Bacillus subtilis)	77	62	294
211	5	1908	2084	gi|410132	ORFX8 (Bacillus subtilis)	77	47	177
217	5	3478	4416	gi|496254	(fibronectin/fibrinogen-binding protein (Streptococcus pyogenes)	77	54	939
232	1	267	998	gi|1407784	orf-1; novel antigen [Staphylococcus aureus]	77	57	732
233	2	1819	1346	gi|467408	unknown [Bacillus subtilis]	77	61	474
243	3	2661	2299	gi|516155	unconventional myosin [Sus scrofa]	77	32	363
299	1	68	769	gi|467436	unknown (Bacillus subtilis]	77	54	702
301	4	1468	1283	gi|950071	ATP-bind. pyrimidine kinase (Mycoplasma capricolum) pir|S48605|S48605 hypothetical protein - Mycoplasma capricolum SGC3) (fragment)	77	48	186
302	5	2741	3211	gi|508980	pheB (Bacillus subtilis)	77	57	471
302	7	3835	4863	gi|147783	ruvB protein [Escherichia coli]	77	60	1029
307	9	5402	4797	gi|1070015	protein-dependent (Bacillus subtilis)	77	60	606
312	1	99	1391	gi|143165	malic enzyme (EC 1.1.1.38) [Bacillus stearothermophilus] pir|A33307|DEBSXS malate dehydrogenase oxaloacetate-decarboxylating) (EC 1.1.1.38) - Bacillus tearothermophilus	77	62	1293
312	2	1541	2443	gi|1399855	(carboxyltransferase beta subunit (Synechococcus PCC7942]	77	58	903
321	5	5666	4595	gi|39844	(fumarase (citG) (aa 1-462) [Bacillus subtilis]	77	65	1071
154	1	47	568	gi|154634	YmoB (Bacillus subtilis]	77	57	522
365	1	2	1021	gi|143374	phosphoribosyl glycinamide synthetase (PUR-D; gtg start codon) Bacillus subtilis)	77	62	1020
374	1	1	708	gi|1405446	transketolase [Bacillus subtilis]	77	61	708
385	1	1128	565	gi|533099	endonuclease III [Bacillus subtilis]	77	63	564
392	2	534	1940	gi|555014	UDP-N-acetyl muramate-alanine ligase (Bacillus subtilis) sp|P40078|MURC_BACSU UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC .3.2.8) (UDP-N- ACETYLMURANOYL-L-ALANINE SYNTHETASE).	77	65	1347
405	5	4079	3570	gi|1303912	YqhW (Bacillus subtilis]	77	64	510
487	4	1302	1472	gi|433427	ORF1 gene product [Acinetobacter calcoaceticus]	77	48	171
522	1	2	562	pir|A01179|SYNS	[tyrosine--tRNA ligase (EC 6.1.1.1) - Bacillus stearothermophilus	77	63	561

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
523	2	1587	1351	gi|1387979	44% identity over 302 residues with hypothetical protein from Synechocystis sp. accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferase; two potential membrane-spanning helices [Bacillus subti]	77	48	237
536	2	983	612	gi|143366	adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir|C29326|N2BSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	77	61	372
548	2	339	872	gi|143387	aspartate transcarbamylase [Bacillus subtilis]	77	56	534
597	1	2	481	gi|904198	hypothetical protein [Bacillus subtilis]	77	33	480
633	2	1747	1313	gi|387577	ORF1A [Bacillus subtilis]	77	64	435
642	1	85	360	gi|46971	epiP gene product [Staphylococcus epidermidis]	77	61	276
659	1	125	1219	gi|1072381	glutamyl-aminopeptidase [Lactococcus lactis]	77	62	1095
670	4	1587	1820	gi|1122760	unknown [Bacillus subtilis]	77	58	234
789	1	2	391	gi|1377823	aminopeptidase [Bacillus subtilis]	77	65	390
815	1	10	573	gi|1303861	yqgN [Bacillus subtilis]	77	49	564
890	1	1	225	gi|1204844	H. influenzae predicted coding region H10594 [Haemophilus influenzae]	77	55	225
1083	1	3	188	gi|460828	B969 [Saccharomyces cerevisiae]	77	66	186
1942	1	415	209	gi|160047	p101/acidic basic repeat antigen (Plasmodium falciparum) pir|A29232|A29232 101K malaria antigen precursor - Plasmodium alciparum (strain Camp]	77	38	207
2559	1	1	171	gi|499034	M. jannaschii predicted coding region MJ0255 [Methanococcus jannaschii]	77	61	171
2933	2	243	401	gi|42370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir|S01788|S01788 formate C-acetyltransferase (EC 2.3.1.54) - cherichia coli	77	72	159
2966	1	56	292	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	77	45	237
2976	1	614	309	gi|140003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p|P21129|ODO1_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGENASE).	77	60	306
2979	2	678	400	gi|1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	77	61	279
2988	1	601	377	gi|438465	Probable operon with orfP. Possible alternative initiation codon, aaas 2151-2153. Homology with acetyltransferases.: putative Bacillus subtilis]	77	55	225
2990	1	331	167	gi|142562	ATP synthase epsilon subunit [Bacillus megaterium] pir|B28599|PWBSEH H+-transporting ATP synthase (EC 3.6.1.34) psilon chain - Bacillus megaterium	77	63	165
3012	1	3	389	gi|488410	alcohol dehydrogenase 2 [Entamoeba histolytica]	77	56	387
3057	1	1	195	gi|468764	mocR gene product [Rhizobium meliloti]	77	50	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	start (nt)	stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
4008	1	726	400	gi|603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 HutI protein, imidazolone-5-propionate hydrolase Bacillus subtilis]	77	52	327
4048	1	703	386	gi|216278	gramicidin S synthetase 1 [Bacillus brevis]	77	55	318
4110	1	3	368	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	366
4115	1	1	348	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	77	65	348
4225	1	590	297	gi|1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	77	60	294
4611	2	494	327	gi|508979	GTP-binding protein [Bacillus subtilis]	77	57	168
4668	1	361	182	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	180
25	1	2	1627	gi|1150620	HmsA [Streptococcus pneumoniae]	76	58	1626
38	5	1488	2537	pir|A43577|A435	regulatory protein pforR - Clostridium perfringens	76	57	1050
52	5	2962	4041	gi|1161061	dioxygenase [Methylobacterium extorquens]	76	62	1080
56	20	27389	27955	gi|467402	unknown [Bacillus subtilis]	76	56	567
57	15	12046	12219	gi|1206040	weak similarity to keratin [Caenorhabditis elegans]	76	40	174
91	2	1062	2261	gi|475715	acetyl coenzyme A acetyltransferase (thiolase) [Clostridium cetobutylicum]	76	57	1200
98	2	818	1624	gi|467422	unknown [Bacillus subtilis]	76	62	807
98	5	2965	3228	gi|897793	y98 gene product [Pediococcus acidilactici]	76	52	264
98	8	5922	6326	gi|467427	methionyl-tRNA synthetase [Bacillus subtilis]	76	53	405
104	3	1322	1885	gi|216151	DNA polymerase (gene L: ttg start codon) [Bacteriophage SPO2] gi|579197 SPO2 DNA polymerase (aa 1-648) [Bacteriophage SPO2] pir|A21498|DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	76	63	564
124	9	8134	7055	gi|853776	peptide chain release factor 1 [Bacillus subtilis] pir|S55437|S55437 peptide chain release factor 1 - Bacillus subtilis	76	58	1080
164	5	2812	3311	gi|1204976	prolyl-tRNA synthetase [Haemophilus influenzae]	76	53	480
168	2	2617	1841	gi|1177253	putative ATP-binding protein of ABC-type [Bacillus subtilis]	76	58	777
189	2	163	888	gi|467384	unknown [Bacillus subtilis]	76	63	726
235	3	2253	3518	gi|142936	folyl-polyglutamate synthetase [Bacillus subtilis] pir|B40646|B40646 folC - Bacillus subtilis	76	53	1266
236	1	335	925	gi|1146197	putative [Bacillus subtilis]	76	54	591
237	8	5323	5541	gi|1279261	F13G3.6 [Caenorhabditis elegans]	76	47	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
263	5	5490	4585	gi|1510348	dihydrodipicolinate synthase [Methanococcus jannaschii]	76	49	906
304	3	1051	1794	gi|666982	putative membrane spanning subunit [Bacillus subtilis] pir|S52382|S52382 probable membrane spanning protein - Bacillus subtilis	76	60	744
312	4	3611	4624	gi|143312	6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) [Bacillus tearothermophilus]	76	56	1014
343	1	2	1036	gi|405956	yveE [Escherichia coli]	76	59'	1035
347	1	409	1701	gi|396304	acetylornithine deacetylase [Escherichia coli]	76	72	1293
358	1	672	1907	gi|1146215	39.0% identity to the Escherichia coli S1 ribosomal protein; putative [Bacillus subtilis]	76	58	1236
371	1	1	222	gi|537084	alternate gene name mgt; CG Site No. 497 [Escherichia coli] pir|S56468|S56468 mgtA protein - Escherichia coli	76	61	222
379	4	4331	4858	gi|143268	dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) [Bacillus subtilis]	76	61	528
404	5	4022	4492	gi|1303823	YqtG [Bacillus subtilis]	76	60	471
411	1	2	307	gi|486025	ORF YKL027w [Saccharomyces cerevisiae]	76	55	306
472	3	4356	2854	gi|1405464	AlsT [Bacillus subtilis]	76	57	1503
546	1	273	995	gi|153821	streptococcal pyrogenic exotoxin type C (speC) precursor Streptococcus pyogenes]	76	36	723
588	1	1054	557	gi|1002520	MutS [Bacillus subtilis]	76	61	498
591	1	16	735	gi|885934	ClpB [Synechococcus sp.]	76	44	720
602	2	175	798	gi|1486422	OppD homologue [Rhizobium sp.]	76	52	624
619	2	547	290	gi|330613	major capsid protein [Human cytomegalovirus]	76	47	258
660	4	2568	3302	gi|904199	hypothetical protein [Bacillus subtilis]	76	55	735
677	1	452	228	gi|40177	spo0F gene product [Bacillus subtilis]	76	58	225
962	1	24	206	gi|142443	adenylosuccinate synthetase [Bacillus subtilis] sp|P29726|PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP-ASPARTATE LIGASE).	76	67	183
978	1	1158	580	gi|1511333	M. jannaschii predicted coding region MJ1322 [Methanococcus jannaschii]	76	56	579
997	1	486	244	gi|1467154	No definition line found [Mycobacterium leprae]	76	38	243
1563	1	529	266	gi|1301984	YqkG [Bacillus subtilis]	76	52	264
2184	1	361	182	gi|506706	CapJ [Staphylococcus aureus]	76	38	180
2572	1	1	387	gi|153898	transport protein [Salmonella typhimurium]	76	65	387

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2942	1	29	400	gi|710020	nitrite reductase [nirB] [Bacillus subtilis]	76	59	372
2957	1	377	216	gi|1511251	hypothetical protein (SP:P42404) [Methanococcus jannaschii]	76	47	162
2980	1	554	279	gi|1405464	AlsT [Bacillus subtilis]	76	53	276
3015	1	649	326	gi|408115	ornithine acetyltransferase [Bacillus subtilis]	76	61	324
3124	1	13	174	gi|882705	ORF_o401 [Escherichia coli]	76	65'	162
3179	1	1	161	gi|1164477	[ferredoxin-dependent glutamate synthase [Zea mays] pir|A38596|A38596 glutamate synthase (ferredoxin) (EC 1.4.7.1) - site	76	53	159
3789	1	2	379	gi|139956	IIGlc [Bacillus subtilis]	76	55	378
3892	1	3	314	gi|1510198	[ferripyochelin binding protein [Methanococcus jannaschii]	76	52	312
3928	1	798	400	gi|143016	permease [Bacillus subtilis]	76	59	399
4159	1	757	386	sp|P80544|MRSP_	METHICILLIN-RESISTANT SURFACE PROTEIN (FRAGMENTS).	76	66	372
4204	1	17	331	gi|1296464	ATPase [Lactococcus lactis]	76	56	315
4199	1	494	249	gi|987255	Menkes disease gene [Homo sapiens]	76	48	246
4506	1	2	313	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	76	47	312
4546	1	477	247	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	76	61	231
4596	1	379	191	gi|550027	cellulose synthase [Acetobacter xylinum]	76	70	189
4	5	5257	4337	gi|882532	ORF_o294 [Escherichia coli]	75	59	921
6	1	164	952	gi|40960	OTCase [Escherichia coli]	75	56	789
12	3	5935	3944	gi|467336	unknown [Bacillus subtilis]	75	57	1992
23	1R	18272	17310	gi|1296433	O-acetylserine sulfhydrylase B [Alcaligenes eutrophus]	75	55	963
25	3	2356	3393	gi|1502419	PlsX [Bacillus subtilis]	75	56	1038
36	8	5765	6037	gi|1256517	unknown [Schizosaccharomyces pombe]	75	45	273
46	13	11186	12058	gi|48972	nitrate transporter [Synechococcus sp.]	75	46	873
51	7	3474	3677	gi|143607	sporulation protein [Bacillus subtilis]	75	61	204
53	16	16850	16590 '	gi|143402	recombination protein (ttg start codon). [Bacillus subtilis] gi|1301923 RecN [Bacillus subtilis]	75	51	261
74	3	3572	2568	gi|1204847	ornithine carbamoyltransferase [Haemophilus influenzae]	75	61	1005

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
85	3	4628	3930	gi|143368	phosphoribosylformyl glycinamidine synthetase I (PUR-L; gtg start codon) [Bacillus subtilis]	75	63	699
85	5	5588	4878	gi|143367	phosphoribosyl aminoidazole succinocarboxamide synthetase (PUR-C; tg start codon) [Bacillus subtilis]	75	55	711
85	8	6625	7530	gi|303916	YqlA [Bacillus subtilis]	75	53	906
87	3	2340	3590	gi|1064813	homologous to sp:PHOR_BACSU [Bacillus subtilis]	75	56	1251
87	6	6084	6896	gi|1064810	function unknown [Bacillus subtilis]	75	61	813
108	2	1844	1503	gi|1001834	hypothetical protein [Synechocystis sp.]	75	51	342
110	3	1748	3727	gi|1147593	putative ppGpp synthetase [Streptomyces coelicolor]	75	55	1980
110	7	4153	5252	gi|1177251	clwD gene product [Bacillus subtilis]	75	75	900
120	14	11266	10649	gi|1524394	ORF-2 upstream of gbsAB operon [Bacillus subtilis]	75	55	618
121	5	2050	4221	gi|1154632	NrdE [Bacillus subtilis]	75	54	2172
124	1	283	143	gi|405622	unknown [Bacillus subtilis]	75	56	141
128	1	81	1139	gi|143316	[gnp] gene products [Bacillus megaterium]	75	48	1059
130	8	5760	5903	gi|1256654	54.8% identity with Neisseria gonorrhoeae regulatory protein PilB; putative [Bacillus subtilis]	75	62	144
136	2	4480	3185	gi|467403	seryl-tRNA synthetase [Bacillus subtilis]	75	54	1296
161	10	5419	5798	gi|1001195	hypothetical protein [Synechocystis sp.]	75	55	360
172	4	3819	2995	gi|755153	ATP-binding protein [Bacillus subtilis]	75	52	825
179	1	2024	1107	gi|143037	porphobilinogen deaminase [Bacillus subtilis]	75	58	918
195	10	9529	9374	su|125745|YCFN_	HYPOTHETICAL PROTEIN IN PURB 5'REGION (ORF-15) (FRAGMENT)	75	60	156
200	4	2605	4596	gi|242440	ATP-dependent nuclease [Bacillus subtilis]	75	56	1992
206	3	6900	5620	gi|1256135	YbbF [Bacillus subtilis]	75	53	1281
216	2	159	389	gi|1052800	unknown [Schizosaccharomyces pombe]	75	58	231
229	1	29	847	gi|1205958	branched chain aa transport system II carrier protein [Haemophilus influenzae]	75	49	819
230	2	518	1714	gi|971337	nitrite extrusion protein [Bacillus subtilis]	75	53	1197
231	1	2240	1122	gi|1002521	MutL [Bacillus subtilis]	75	54	1119
233	3	1314	1859	gi|467405	unknown [Bacillus subtilis]	75	59	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
269	1	325	164	gi|1511246	methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]	75	50	162
282	1	1189	772	gi|1511604	M. jannaschii predicted coding region MJ1651 [Methanococcus jannaschii]	75	46	618
304	4	1773	2261	gi|1205328	surfactin [Haemophilus influenzae]	75	55	489
312	3	2437	3387	gi|285621	undefined open reading frame [Bacillus stearothermophilus]	75	62	951
312	5	4622	6403	gi|1041097	Pyruvate Kinase [Bacillus psychrophilus]	75	57*	1782
319	1	351	877	gi|1212728	yqhI [Bacillus subtilis]	75	54	525
320	5	4321	5031	gi|1070361	OMP decarboxylase [Lactococcus lactis]	75	56	711
320	6	5010	5642	gi|143394	OMP-PRPP transferase [Bacillus subtilis]	75	60	633
337	4	1519	2088	gi|487433	citrate synthase II [Bacillus subtilis]	75	58	570
394	2	669	1271	gi|304976	matches PS00017; ATP_GTP_A and PS00301; EFACTOR_GTP; similar to longation [factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli]	75	51	603
423	1	127	570	gi|1183839	unknown [Pseudomonas aeruginosa]	75	59	444
433	2	1603	1929	gi|149211	acetolactate synthase [Klebsiella pneumoniae]	75	63	327
446	2	176	1540	gi|312441	dihydroorotase [Bacillus caldolyticus]	75	62	1365
486	1	494	249	gi|1149682	potF gene product [Clostridium perfringens]	75	55	246
496	1	3	794	gi|143582	spoIIIEA protein [Bacillus subtilis]	75	59	792
498	2	824	1504	gi|143328	phoP protein (put.); putative [Bacillus subtilis]	75	47	681
499	2	1061	1624	gi|1387979	44% identity over 302 residues with hypothetical protein from Synechocystis sp. accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferase; two potential membrane-spanning helices [Bacillus subti]	75	51	564
568	1	641	453	pir|JC4110|JC4	triacylglycerol lipase (EC 3.1.1.3) 2 - Mycoplasma mycoides subsp. mycoides [SGC3]	75	50	189
613	2	430	233	gi|330993	tegument protein [Saimirine herpesvirus 2]	75	75	198
621	1	1	525	gi|529754	spec [Streptococcus pyogenes]	75	43	525
642	5	1809	2474	gi|1176401	EpiG [Staphylococcus epidermidis]	75	51	666
646	2	454	657	gi|172442	ribonuclease P [Saccharomyces cerevisiae]	75	37	204
657	1	3	347	gi|882541	ORF_o236 [Escherichia coli]	75	47	345
750	1	1662	832	gi|46971	epiP gene product [Staphylococcus epidermidis]	75	57	831

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
754	1	2	481	gi|1303901	YqhT [Bacillus subtilis]	75	57	480
763	2	563	393	gi|1205145	multidrug resistance protein [Haemophilus influenzae]	75	51	171
775	1	961	482	pir|B36889|D368	leuA protein, inactive - Lactococcus lactis subsp. lactis (strain IL140)	75	63	480
793	1	1	180	gi|143316	[gap] gene products [Bacillus megaterium]	75	57	180
800	1	318	160	gi|509411	NPRA protein [Azorhizobium caulinodans]	75	34*	159
811	1	1117	560	gi|143434	Rho factor [Bacillus subtilis]	75	60	558
940	1	493	329	gi|1276985	arginase [Bacillus caldovelox]	75	50	165
971	2	37	252	gi|1001373	hypothetical protein [Synechocystis sp.]	75	58	216
1059	1	384	232	gi|726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]	75	67	153
1109	2	219	374	gi|143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-).	75	53	156
126A	1	271	137	gi|104135	ornithine acetyltransferase [Bacillus stearothermophilus] sp|Q07908|ARGJ_BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACETYLTRANSFERASE (ORNITHINE TRANSACETYLASE) (OATASE) / AMINO-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-ACETYLGLUTAMATE YNTHA	75	63	135
1500	1	324	163	gi|1205488	excinuclease ABC subunit B [Haemophilus influenzae]	75	57	162
1529	1	79A	400	gi|1002521	HutL [Bacillus subtilis]	75	54	199
3010	1	770	387	gi|1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]	75	54	384
3105	1	1	180	gi|104107	Pyruvate Kinase [Bacillus psychrophilus]	75	57	180
3117	1	45	212	gi|899117	peptide synthetase module [Microcystis aeruginosa] pir|S49111|S49111 probable amino acid activating domain - lcrocystis aeruginosa (fragment) (SUB 144-528)	75	42	168
3139	2	139	345	gi|145294	adenine phosphoribosyl-transferase [Escherichia coli]	75	66	207
3880	1	618	310	gi|1009166	Respiratory nitrate reductase [Bacillus subtilis]	75	58	309
3911	1	48	401	gi|433991	ATP synthase subunit beta [Bacillus subtilis]	75	68	354
3957	1	2	379	pir|D16AH9|D368	3-isopropylmalate dehydratase (EC 4.2.1.33) chain leuC - Lactococcus lactis subsp. lactis (strain IL140)	75	65	378
4005	1	5	259	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	75	48	255
4080	1	73	333	gi|415855	deoxyribose aldolase [Mycoplasma hominis]	75	59	261

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4111	1	1	339	gi|149435	putative [Lactococcus lactis]	75	57	339
4136	1	602	303	gi|1450648	hsdM gene of Ecoprrl gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	75	56	300
4144	1	668	336	gi|48972	nitrate transporter [Synechococcus sp.]	75	49	333
4237	1	664	374	gi|1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	75	55'	291
4306	2	73	318	gi|294260	major surface glycoprotein [Pneumocystis carinii]	75	68	246
4343	1	715	359	gi|1204652	methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]	75	52	357
4552	1	620	312	gi|296464	ATPase [Lactococcus lactis]	75	55	309
38	9	5776	6126	gi|443793	NupC [Escherichia coli]	74	50	351
50	8	6910	6221	gi|1239988	hypothetical protein [Bacillus subtilis]	74	55	690
56	9	10770	12221	gi|1000451	TreP [Bacillus subtilis]	74	57	1452
64	2	2266	1622	gi|41015	aspartate-tRNA ligase [Escherichia coli]	74	57	645
66	6	5063	4848	gi|1212729	YqhJ [Bacillus subtilis]	74	47	216
67	18	14314	14897	gi|1510631	endoglucanase [Methanococcus jannaschii]	74	52	564
102	15	12561	13136	gi|149429	putative [Lactococcus lactis]	74	67	576
102	16	13121	14419	gi|149435	putative [Lactococcus lactis]	74	57	1299
108	4	4873	3902	gi|39478	ATP binding protein of transport ATPases [Bacillus firmus] tr|S15486|S15486 ATP-binding protein - Bacillus firmus p|P26946|YATR_BACFI HYPOTHETICAL ATP-BINDING TRANSPORT PROTEIN.	74	59	972
116	5	8574	7093	gi|1205430	dipeptide transport system permease protein [Haemophilus influenzae]	74	49	1482
120	7	4342	4803	gi|146970	ribonucleoside triphosphate reductase [Escherichia coli] pir|A47331|A47331 anaerobic ribonucleotide reductase - Escherichia oil	74	58	462
121	7	5961	6581	gi|1107528	ttg start [Campylobacter coli]	74	51	621
128	3	2320	3531	gi|143318	phosphoglycerate kinase [Bacillus megaterium]	74	57	1212
130	7	5237	5791	gi|1256653	DNA-binding protein [Bacillus subtilis]	74	60	555
136	3	6745	5150	gi|143076	histidase [Bacillus subtilis]	74	58	1596
145	2	664	1368	gi|407773	devA gene product [Anabaena sp.]	74	45	705
152	1	552	277	gi|1377833	unknown [Bacillus subtilis]	74	54	276

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
164	10	11064	11375	gi|580900	[ORF3 gene product (Bacillus subtilis)]	74	52	312
175	2	3109	2624	gi|642656	[unknown (Rhizobium meliloti)]	74	34	486
175	9	6064	5612	gi|854656	[Na/H antiporter system ORF2 (Bacillus alcalophilus)]	74	46	453
195	11	11346	10339	gi|1204430	[hypothetical protein (SP:P25745) (Haemophilus influenzae)]	74	55	1008
205	17	9619	9059	gi|1044979	[ribosomal protein L6 (Bacillus subtilis)]	74	64	561
236	7	5574	6710	gi|1146207	[putative (Bacillus subtilis)]	74	63	1137
241	3	4521	3334	gi|694121	[malate thiokinase (Methylobacterium extorquens)]	74	52	1188
246	6	3305	2799	gi|467374	[single strand DNA binding protein (Bacillus subtilis)]	74	64	507
249	4	6551	5313	gi|1524397	[glycine betaine transporter OpuD (Bacillus subtilis)]	74	55	1239
261	7	4389	4081	gi|809542	[CbrB protein (Erwinia chrysanthemi)]	74	42	309
278	6	5714	4665	gi|1204872	[ATP-binding protein (Haemophilus influenzae)]	74	54	1050
309	1	1220	666	gi|1205579	[hypothetical protein (GB:U14003_103) (Haemophilus influenzae)]	74	53	555
315	2	1473	862	gi|143398	[quinol oxidase (Bacillus subtilis)]	74	57	612
320	1	1	1065	gi|143389	[glutaminase of carbamyl phosphate synthetase (Bacillus subtilis) pir|E39845|E39845 carbamoyl-phosphate synthase glutamine-hydrolyzing) (EC 6.3.5.5), pyrimidine-repressible, small chain - Bacillus subtilis]	74	60	1065
380	2	382	1128	gi|534857	[ATPase subunit a (Bacillus stearothermophilus)]	74	56	747
405	2	1742	1311	gi|1303915	[Yqh2 (Bacillus subtilis)]	74	65	432
433	5	2503	3270	gi|473902	[alpha-acetolactate synthase (Lactococcus lactis)]	74	56	768
452	1	1	942	gi|413982	[lpa-58r gene product (Bacillus subtilis)]	74	52	942
461	1	3	1193	gi|558494	[homoserine dehydrogenase (Bacillus subtilis)]	74	51	1191
461	2	1174	1407	gi|40211	[threonine synthase (thrC) (AA 1-252) (Bacillus subtilis) lr|A25364|A25364 threonine synthase (EC 4.2.99.2) - Bacillus subtilis]	74	56	234
462	2	402	734	gi|142520	[thioredoxin (Bacillus subtilis)]	74	62	333
478	1	574	320	gi|1499005	[glycyl-tRNA synthetase (Methanococcus jannaschii)]	74	52	255
501	2	739	1740	gi|217040	[acid glycoprotein (Streptococcus pyogenes)]	74	58	1002
551	2	4083	2791	gi|143040	[glutamate-1-semialdehyde 2,1-aminotransferase (Bacillus subtilis) pir|D42728|D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC 4.3.8) - Bacillus subtilis]	74	51	1293

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Acession	match gene name	% sim	% ident	length (nt)
573	1	1	477	gi|1006605	hypothetical protein (Synechocystis sp.)	74	45	477
596	2	1780	1298	gi|1303853	YqgF [Bacillus subtilis]	74	55	483
618	2	2924	1758	gi|1146237	21.4% of identity to trans-acting transcription factor of Saccharomyces cerevisiae; 25% of identity to sucrose synthase of Zea mays; putative [Bacillus subtilis]	74	55	1167
659	2	1269	1595	gi|1072380	ORF3 [Lactococcus lactis]	74	62	327
724	1	373	188	gi|143374	phosphoribosyl glycinamide synthetase (PUR-D; gcg start codon) Bacillus subtilis	74	58	186
743	2	604	1209	gi|153833	ORF1; putative [Streptococcus parasanguis]	74	50	606
836	1	2	259	gi|143458	ORF V [Bacillus subtilis]	74	47	258
989	2	443	724	gi|1303994	YqxM [Bacillus subtilis]	74	46	282
1106	1	1	492	gi|46970	epiD gene product [Staphylococcus epidermidis]	74	54	492
1135	2	373	528	gi|141948	ipa-24d gene product [Bacillus subtilis]	74	48	156
1234	1	817	452	gi|495245	rocJ gene product [Erwinia chrysanthemi]	74	36	366
2586	1	2	238	gi|1149701	sbcC gene product [Clostridium perfringens]	74	62	237
2959	1	798	400	gi|1405454	aconitase [Bacillus subtilis]	74	60	399
2962	1	650	363	gi|450686	3-phosphoglycerate kinase [Thermotoga maritima]	74	58	288
2983	1	3	191	gi|1303893	YqhL [Bacillus subtilis]	74	56	189
3018	1	2	223	gi|143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir|D42728|D42728 glutamate-1-semialdehyde 2,1-aminomutase (EC .4.3.8) - Bacillus subtilis	74	56	222
3038	1	510	256	pir|S52915|S52915	nitrate reductase alpha chain - Bacillus subtilis (fragment)	74	57	255
3062	1	374	189	gi|1107528	ttg start [Campylobacter coli]	74	51	186
4035	1	184	360	gi|1022725	unknown [Staphylococcus haemolyticus]	74	64	177
4045	1	607	305	gi|1510977	M. jannaschii predicted coding region MJ0938 [Methanococcus jannaschii]	74	41	303
4283	1	471	304	gi|520844	orf4 [Bacillus subtilis]	74	58	168
4449	1	3	221	gi|580910	peptide-synthetase ORF1 [Bacillus subtilis]	74	54	219
4587	1	458	231	gi|1370207	orf6 [Lactobacillus sake]	74	59	228

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4603	1	29	214	gi|146208	glutamate synthase large subunit (EC 2.6.1.5) [Escherichia coli] pir[A29617|A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli	74	60	186
4670	1	366	184	gi|256135	YbbF [Bacillus subtilis]	74	61	183
5	10	7953	7162	gi|143727	putative [Bacillus subtilis]	73	42	792
11	2	2454	1372	gi|166338	dihydroorotate dehydrogenase [Agrocybe aegerita]	73	55	1083
14	1	2024	1020	gi|143373	phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-HJ) [Bacillus subtilis]	73	54	1005
23	5	5426	4635	gi|1468939	meso-2,3-butanediol dehydrogenase (D-acetoin forming) [Klebsiella pneumoniae]	73	58	792
23	17	17379	16360	gi|297060	ornithine cyclodeaminase [Rhizobium meliloti]	73	37	1020
29	2	692	1273	gi|467442	stage V sporulation [Bacillus subtilis]	73	54	582
31	5	6467	4914	gi|414000	ipa-76d gene product [Bacillus subtilis]	73	55	1554
37	8	8658	7402	gi|429259	pagT gene product [Bacillus subtilis]	73	59	1257
37	9	7738	7562	gi|168367	alpha-isopropylmalate isomerase (put.); putative [Rhizomucor ircinelloides]	73	52	177
38	7	3931	4896	gi|405885	yaiN [Escherichia coli]	73	58	966
44	6	5041	4238	gi|5R0895	unknown [Bacillus subtilis]	73	53	804
44	11	7767	8306	gi|142009	mraD gene product [Escherichia coli]	73	50	540
45	3	2439	3080	gi|1109685	ProW [Bacillus subtilis]	73	47	642
54	13	14036	13794	gi|413931	ipa-7d gene product [Bacillus subtilis]	73	61	243
59	4	1430	2248	gi|147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	73	53	819
65	1	1458	730	gi|677944	AppF [Bacillus subtilis]	73	56	729
80	2	1375	860	gi|580932	murD gene product [Bacillus subtilis]	73	53	516
102	13	10124	11179	gi|580891	3-isopropylmalate dehydrogenase (AA 1 - 365) [Bacillus subtilis] pir[A26522|A26522 3-isopropylmalate dehydrogenase (EC 1.1.1.85) - acillus subtilis	73	55	1056
109	2	3493	2600	gi|1510849	M. jannaschii predicted coding region MJ0775 [Methanococcus jannaschii]	73	40	894
120	8	4782	5756	gi|146970	ribonucleoside triphosphate reductase [Escherichia coli] pir[A47331|A47331 anaerobic ribonucleotide reductase - Escherichia oli	73	56	975
120	9	5726	6223	gi|1204333	anaerobic ribonucleoside-triphosphate reductase [Haemophilus influenzae]	73	62	498

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
132	5	4151	4363	gi|871048	HPSK2 - heavy chain potential motor protein (Giardia intestinalis)	73	43	213
140	6	5952	4324	gi|634107	kdpB (Escherichia coli)	73	59	1629
142	6	7060	5939	gi|410125	rluG gene product (Bacillus subtilis)	73	57	1122
149	4	1866	1717	gi|1460892	heparin binding protein-44, HBP-44 (mice, Peptide, 360 aa) pir|JX0281|JX0281 heparin-binding protein-44 precursor - mouse gi|220434 ORF (Mus musculus) (SUB 2-360)	73	53	150
158	1	1	1431	gi|882504	ORF_f560 (Escherichia coli)	73	57	1431
174	6	5352	4525	gi|1146240	ketopantoate hydroxymethyltransferase (Bacillus subtilis)	73	55	828
175	8	5537	5178	gi|854657	Na/H antiporter system ORF3 (Bacillus alcalophilus)	73	56	360
186	5	6593	5493	gi|467477	unknown (Bacillus subtilis)	73	48	1101
249	6	6283	5729	gi|1524397	glycine betaine transporter OpuD (Bacillus subtilis)	73	56	555
265	4	1873	2280	gi|139848	U3 (Bacillus subtilis)	73	41	408
270	1	328	582	gi|7R0461	220 kDa polyprotein (African swine fever virus)	73	53	255
278	4	4283	3618	gi|1204965	hypothetical 23.3 kd protein (Escherichia coli)	73	49	666
279	3	4984	3593	gi|1185288	isochorismate synthase (Bacillus subtilis)	73	58	1392
291	4	1207	1575	gi|1511440	glutamine--fructose-6-phosphate transaminase (Methanococcus jannaschii)	73	63	369
299	2	735	1166	gi|467437	unknown (Bacillus subtilis)	73	58	432
299	5	2050	3234	gi|467439	temperature sensitive cell division (Bacillus subtilis)	73	53	1185
334	1	1237	728	gi|536655	ORF YBR244w (Saccharomyces cerevisiae)	73	43	510
336	2	1827	1036	gi|790943	urea amidolyase (Bacillus subtilis)	73	51	792
374	3	1389	1874	gi|1405451	YneJ (Bacillus subtilis)	73	55	486
433	4	1916	2554	gi|473902	alpha-acetolactate synthase (Lactococcus lactis)	73	54	639
509	2	1795	1028	gi|467483	unknown (Bacillus subtilis)	73	56	768
513	1	1709	918	gi|1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)	73	56	792
533	2	239	733	gi|1510605	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	73	44	495
546	2	1148	2815	gi|41748	hadM protein (AA 1-520) (Escherichia coli)	73	52	1668
549	1	762	382	gi|1314847	CinA (Bacillus subtilis)	73	57	381
567	1	1346	675	gi|410137	ORFX13 (Bacillus subtilis)	73	58	672

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
716	2	654	1112	gi|1256623	exodeoxyribonuclease (Bacillus subtilis)	73	56	459
772	1	3	677	gi|142010	Shows 70.2% similarity and 48.6% identity to the EnvH protein of almonella typhimurium (Anabaena sp.)	73	57	675
774	1	3	209	gi|409286	bmrU (Bacillus subtilis)	73	52	207
782	1	1	402	gi|143320	[gap] gene products (Bacillus megaterium)	73	56	402
789	2	451	762	gi|1063246	low homology to P14 protein of Heamophilus influenzar and 14.2 kDa protein of Escherichia coli (Bacillus subtilis)	73	56	312
796	1	3	911	gi|853754	ABC transporter (Bacillus subtilis)	73	58	909
806	3	1209	949	gi|143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Bacillus subtilis) pir|JT0481|YWBS tryptophan-tRNA ligase (EC 6.1.1.2) - Bacillus ubtilis	73	51	261
816	2	4839	3097	gi|41748	hsdM protein (AA 1-520) [Escherichia coli]	73	52	1743
839	1	798	400	gi|886906	argininosuccinate synthetase (Streptomyces clavuligerus) pir|S57659|S57659 argininosuccinate synthase (EC 6.3.4.5) - treptomyces clavuligerus	73	59	399
857	1	1	290	gi|348052	acetoin utilization protein (Bacillus subtilis)	73	50	288
1008	1	790	398	gi|40100	rodC (tagl) polypeptide (AA 1-746) [Bacillus subtilis] ir|S06049|S06049 rodC protein - Bacillus subtilis p|P13485|TAGF_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN F.	73	41	393
1018	1	1	213	gi|529357	No definition line found (Caenorhabditis elegans) sp|P46975|STT3_CAEEL OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT ONOLOG.	73	53	213
1011	1	3	491	gi|142706	comG1 gene product [Bacillus subtilus]	73	51	489
1174	1	395	204	gi|149513	alpha3a subunit of laminin 5 (Homo sapiens)	73	60	192
1175	1	655	329	gi|473817	'ORF' [Escherichia coli]	73	57	327
1187	1	3	209	gi|580870	ipa-37d qoxA gene product (Bacillus subtilis)	73	52	207
1206	1	72	245	gi|144816	formyltetrahydrofolate synthetase (FTHFS) (ttg start codon) (EC 3.4.3) (Moorella thermoacetica)	73	43	174
1454	1	423	241	gi|1213253	unknown [Schizosaccharomyces pombe]	73	53	183
1469	1	517	260	gi|1303787	YqeG [Bacillus subtilis]	73	55	258
1761	1	374	189	gi|9135	Hst26Aa gene product [Drosophila simulans]	73	34	186
1849	1	467	243	gi|162307	DNA topoisomerase II [Trypanosoma cruzi]	73	60	225
2055	2	2	400	gi|159381	P47K protein [Rhodococcus erythropolis]	73	34	399
2556	1	2	244	gi|145925	tecB [Escherichia coli]	73	62	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
2947	2	549	400	gi|1184680	polynucleotide phosphorylase [Bacillus subtilis]	73	51	150
2956	1	746	375	gi|143397	quinol oxidase [Bacillus subtilis]	73	58	372
3037	1	655	329	gi|143091	acetolactate synthase [Bacillus subtilis]	73	55	327
3115	1	385	194	gi|323866	overlapping out-of-phase protein [Eggplant mosaic virus] sp|P20129|V70K_EPMV 70 KD PROTEIN.	73	53	192
3603	2	700	527	gi|149521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	174
3743	1	798	400	gi|450688	hadM gene of Ecoprrl gene product [Escherichia coli] pir|S38437|S38437 hadM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	73	54	399
3752	1	640	359	gi|1324193	unknown [Mycobacterium tuberculosis]	73	59	282
3852	1	2	181	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	73	68	180
3914	1	475	239	pir|S13490|S134	Hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chicken (fragment)	73	53	237
3914	2	570	343	gi|528991	unknown [Bacillus subtilis]	73	38	228
4069	1	2	316	gi|40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p|P23129|ODO1_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGENASE).	73	55	315
4165	1	715	365	gi|143521	glutaryl-CoA dehydrogenase precursor [Mus musculus]	73	48	351
4196	1	1	177	gi|809660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir|S49455|S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acilius subtilis	73	60	177
4202	1	572	378	gi|528991	unknown [Bacillus subtilis]	73	38	195
4314	1	2	193	gi|336797	N-acyl-L-amino acid amidohydrolase [Bacillus stearothermophilus] sp|P37112|AMA_BACST N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC .5.1.14) (AMINOACYLASE).	73	47	192
4393	1	3	261	gi|216267	ORF2 [Bacillus megaterium]	73	48	261
35	2	903	1973	gi|1146196	phosphoglycerate dehydrogenase [Bacillus subtilis]	72	53	1071
38	22	19094	17877	(u|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	72	54	1218
38	23	18134	19162	gi|413968	lpa-44d gene product [Bacillus subtilis]	72	54	1029
44	19	11895	12953	gi|516272	unknown [Bacillus subtilis]	72	49	1059
48	7	6248	7117	gi|43499	pyruvate synthase [Halobacterium halobium]	72	49	870
50	7	6563	5691	gi|1205399	proton glutamate symport protein [Haemophilus influenzae]	72	53	871

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
53	9	10521	9259	gi|1303956	YqjE [Bacillus subtilis]	72	52	1263
56	23	29549	29995	gi|467471	unknown [Bacillus subtilis]	72	47	447
69	4	5298	4123	gi|1354775	pfoS/R [Treponema pallidum]	72	46	1176
69	5	4377	4982	gi|1904198	hypothetical protein [Bacillus subtilis]	72	43	606
73	1	2	856	gi|142997	glycerol uptake facilitator [Bacillus subtilis]	72	59	855
98	13	9371	10258	gi|467435	unknown [Bacillus subtilis]	72	50	888
127	1	1	1593	gi|217144	alanine carrier protein (thermophilic bacterium PS3) pir|A45111|A45111 alanine transport protein - thermophilic acterium PS-3	72	56	1593
131	1	5197	2600	gi|153952	polymerase III polymerase subunit (dnaE) [Salmonella typhimurium] pir|A45915|A45915 DNA-directed DNA polymerase (EC 2.7.7.7) III ipha chain - Salmonella typhimurium	72	53	2598
141	4	1040	1978	gi|1405446	transketolase [Bacillus subtilis]	72	54	939
149	8	2819	2535	gi|606234	secY [Escherichia coli]	72	44	285
149	17	5472	5245	gi|1304472	DNA polymerase [Unidentified phycodnavirus clone O7U6]	72	55	228
154	1	1	210	gi|1205620	ferritin like protein [Haemophilus influenzae]	72	40	210
155	1	2207	1320	gi|391610	farnesyl diphosphate synthase [Bacillus stearothermophilus] pir|JX0257|JX0257 geranyltranstransferase (EC 2.5.1.10) - Bacillus tearothermophilus	72	57	888
180	1	2	328	gi|433630	A180 [Saccharomyces cerevisiae]	72	62	327
184	3	1145	3553	gi|1205110	virulence associated protein homolog [Haemophilus influenzae]	72	49	2409
195	2	1923	1279	gi|1001730	hypothetical protein [Synechocystis sp.]	72	45	645
206	13	14646	15869	gi|1064807	ORNITHINE AMINOTRANSFERASE [Bacillus subtilis]	72	50	1224
209	2	463	932	gi|1204666	hypothetical protein (GB:X73124_53) [Haemophilus influenzae]	72	60	471
215	2	764	522	gi|881513	insulin receptor homolog [Drosophila melanogaster] pir|S57245|S57245 insulin receptor homolog - fruit fly (Drosophila elanogaster) (SUB 46-216)	72	63	243
224	1	2	790	gi|945974	sucrose repressor [Staphylococcus xylosus]	72	54	789
233	1	1526	765	gi|1408493	homologous to SwissProt:YIDA_ECOLI hypothetical protein [Bacillus subtilis]	72	52	762
240	1	220	1485	gi|537049	ORF_d470 [Escherichia coli]	72	52	1266
245	1	3	1340	gi|1204578	hypothetical protein (GB:U06949_1) [Haemophilus influenzae]	72	46	1338

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
255	2	2108	1245	gi|1340128	ORF1 [Staphylococcus aureus]	72	59	864
304	2	285	1094	gi|1205330	glutamine-binding periplasmic protein [Haemophilus influenzae]	72	52	810
307	10	5326	5039	gi|1070015	protein-dependent [Bacillus subtilis]	72	53	288
315	1	517	260	gi|143399	quinol oxidase [Bacillus subtilis]	72	55	258
316	11	9622	9308	gi|1204445	hypothetical protein (SP:P27857) [Haemophilus influenzae]	72	52	315
337	3	926	1609	gi|487433	citrate synthase II [Bacillus subtilis]	72	55	684
364	7	12538	10493	gi|1510643	ferrous iron transport protein B [Methanococcus jannaschii]	72	53	2046
409	2	340	1263	gi|1402944	orfAM1 gene product [Bacillus subtilis]	72	49	924
441	3	2177	1590	gi|1312379	highly conserved among eubacteria (Clostridium acetobutylicum) pir|S34312|S34312 hypothetical protein V - Clostridium cetobutylicum	72	48	588
453	6	2654	2505	pir|S00601|BXSA	antibacterial protein 3 - Staphylococcus haemolyticus	72	70	150
460	1	2	625	gi|1016162	ABC transporter subunit [Cyanophora paradoxa]	72	51	624
463	1	3253	1628	gi|666014	The polymorphysm (RFLP) of this gene is associated with susceptibility to essential hypertension. The SA gene product has light homology to acetyl-CoA synthetase [Homo sapiens]	72	60	1626
480	4	1047	3466	gi|433992	ATP synthase subunit epsilon [Bacillus subtilis]	72	53	420
502	1	1086	586	gi|310859	ORF2 [Synechococcus sp.]	72	50	501
519	1	81	1184	gi|1303704	YrkE [Bacillus subtilis]	72	54	1104
559	1	3	746	gi|1107530	ceuD gene product [Campylobacter coli]	72	56	744
575	1	1142	573	gi|1303866	YqgS [Bacillus subtilis]	72	56	570
671	1	2	592	gi|1204497	protein-export membrane protein [Haemophilus influenzae]	72	44	591
679	2	295	1251	gi|563258	virulence-associated protein E [Dichelobacter nodosus]	72	52	957
687	2	295	957	gi|1146214	44% identical amino acids with the Escherichia coli smba supress; putative [Bacillus subtilis]	72	49	663
837	1	1	435	gi|1146183	putative [Bacillus subtilis]	72	54	435
868	1	150	788	gi|1377842	unknown [Bacillus subtilis]	72	55	639
922	1	130	432	gi|1088269	unknown protein [Azotobacter vinelandii]	72	58	303
941	1	2	238	gi|153929	NADPH-sulfite reductase (flavoprotein component [Salmonella yphimurium]	72	49	237
980	1	840	421	gi|853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	72	59	420

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1209	1	383	213	gi|144735	neurotoxin type B [Clostridium botulinum]	72	44	171
1469	2	671	474	gi|1205458	hypothetical protein (GB:D26562_47) [Haemophilus influenzae]	72	63	198
1956	1	727	365	gi|154409	hexosephosphate transport protein (Salmonella typhimurium) pir|B41853|B41853 hexose phosphate transport system regulatory rotein uhpB - Salmonella typhimurium	72	44	363
2101	1	3	401	gi|1303950	YqiY [Bacillus subtilis]	72	50	399
2503	1	569	399	gi|149713	formate dehydrogenase (Methanobacterium formicicum) pir|A42712|A42712 formate dehydrogenase (EC 1.2.1.2) - ethanobacterium formicicum	72	56	171
2967	1	3	155	gi|1212729	YqhJ [Bacillus subtilis]	72	46	153
3004	1	167	185	gi|1665999	hypothetical protein [Bacillus subtilis]	72	55	183
3109	1	278	141	gi|413968	ipa-44d gene product [Bacillus subtilis]	72	45	138
3171	1	3	287	gi|1515938	glutamate synthase (ferredoxin) [Synechocystis sp.] pir|646957|646957 glutamate synthase (ferredoxin) - ynechocystis sp.	72	52	285
3771	1	26	367	gi|1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	72	63	342
3951	1	1	222	gi|1500409	M. jannaschii predicted coding region MJ1519 [Methanococcus jannaschii]	72	38	222
4190	1	721	362	gi|39956	IIGlc [Bacillus subtilis]	72	57	360
4444	1	3	347	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	72	55	345
6	2	711	1200	gi|537095	ornithine carbamoyltransferase [Escherichia coli]	71	??	270
11	15	11350	10859	gi|532309	25 kDa protein [Escherichia coli]	71	47	492
19	2	1248	2435	gi|1244574	D-alanine:D-alanine ligase [Enterococcus hirae]	71	52	1188
21	2	898	1488	gi|149629	anthranilate synthase component 2 (Leptospira biflexa) pir|C32840|C32840 anthranilate synthase (EC 4.1.3.27) component II Leptospira biflexa	71	45	591
34	1	1	567	gi|1303983	YqkF [Bacillus subtilis]	71	59	567
37	3	3192	2606	gi|1209681	glutamate-rich protein [Bacillus firmus]	71	50	387
38	18	12250	12462	gi|927645	arginyl endopeptidase [Porphyromonas gingivalis]	71	50	213
39	3	1246	4431	pir|S09411|S09411	spoIIIE protein - Bacillus subtilis	71	49	3186
53	14	15770	14760	gi|142611	branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus subtilis]	71	58	1011
54	11	13461	12625	gi|143014	gnt repressor [Bacillus subtilis]	71	46	837

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
57	7	7152	5860	gi|1508175	EIIC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli]	71	48	1293
57	18	13897	14314	gi|1063247	high homology to flavohemoprotein (Haemoglobin-like protein) of Alcaligenes eutrophus and Saccharomyces cerevisiae [Bacillus subtilis]	71	56	438
62	16	9831	10955	gi|1303926	YqiG [Bacillus subtilis]	71	54	1125
70	12	8505	8966	gi|147198	phnE protein [Escherichia coli]	71	38	462
86	5	2394	2089	gi|904205	hypothetical protein [Bacillus subtilis]	71	51	306
96	7	7601	8269	gi|709991	hypothetical protein [Bacillus subtilis]	71	49	669
100	6	4822	5931	gi|1060848	Opine dehydrogenase [Arthrobacter sp.]	71	45	1110
103	1	1062	532	gi|143089	iep protein [Bacillus subtilis]	71	41	531
109	18	15312	15695	gi|413985	ipa-61d gene product [Bacillus subtilis]	71	57	384
113	1	630	316	gi|663254	probable protein kinase [Saccharomyces cerevisiae]	71	57	315
114	5	6598	5603	gi|143156	membrane bound protein [Bacillus subtilis]	71	40	996
133	2	3087	1723	gi|1303913	YqhX [Bacillus subtilis]	71	53	1165
149	19	6335	5895	gi|529650	G40P [Bacteriophage SPP1]	71	51	441
154	5	3635	3087	gi|425488	repressor protein [Streptococcus sobrinus]	71	47	549
164	11	11354	11649	gi|49318	ORF4 gene product [Bacillus subtilis]	71	52	336
169	5	1936	2745	gi|1403403	unknown [Mycobacterium tuberculosis]	71	56	810
193	2	272	1234	gi|1303788	YqeH [Bacillus subtilis]	71	49	963
205	1	1743	895	gi|1215694	GlnQ [Mycoplasma pneumoniae]	71	46	849
233	4	1849	2022	gi|633732	ORF1 [Campylobacter jejuni]	71	50	174
237	7	4501	5169	gi|149384	HisIE [Lactococcus lactis]	71	54	669
272	4	2848	2273	gi|709991	hypothetical protein [Bacillus subtilis]	71	48	576
274	2	618	1496	gi|143035	NAD(P)H:glutamyl-transfer RNA reductase [Bacillus subtilis] pir|A35252|A35252 5-aminolevulinate synthase (EC 2.3.1.37) - acillus subtilis	71	53	879
276	5	3349	2220	gi|1303562	ORF210 [Escherichia coli]	71	50	630
287	1	136	660	gi|310634	20 kDa protein [Streptococcus gordonii]	71	53	525
288	6	3322	2771	gi|1256625	putative [Bacillus subtilis]	71	47	552

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
301	6	3492	2461	gi|467417	similar to lysine decarboxylase [Bacillus subtilis]	71	57	1032
306	4	6607	5222	gi|1256618	transport protein [Bacillus subtilis]	71	56	1386
307	2	1536	925	gi|632683	orfC [Mycoplasma capricolum]	71	45	612
310	5	5793	5146	gi|348052	acetoin utilization protein [Bacillus subtilis]	71	51	648
322	1	2	1303	gi|1001819	hypothetical protein [Synechocystis sp.]	71	46	1302
333	4	4171	3995	gi|467473	unknown [Bacillus subtilis]	71	57	177
350	2	548	922	gi|551879	ORF 1 [Lactococcus lactis]	71	55	375
375	4	1860	3071	gi|467447	unknown [Bacillus subtilis]	71	57	1212
380	5	1560	2102	gi|142557	ATP synthase b subunit [Bacillus megaterium]	71	43	543
414	2	251	637	gi|580904	homologous to E.coli rnpA [Bacillus subtilis]	71	49	387
424	1	335	1354	gi|581305	L-lactate dehydrogenase [Lactobacillus plantarum]	71	57	1020
436	4	3701	3270	pir|PN0501|PN05	phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis (fragment)	71	66	432
482	1	3	1280	gi|410142	ORFX18 [Bacillus subtilis]	71	49	1278
525	3	2272	1844	gi|143370	phosphoribosylpyrophosphate amidotransferase (PUR-F; EC 2.4.2.14) Bacillus subtilis]	71	56	429
529	4	2732	2047	gi|606150	ORF_f309 [Escherichia coli]	71	43	691
563	1	22	969	gi|1237015	ORF4 [Bacillus subtilis]	71	53	948
581	1	506	255	gi|1301730	T2SO3.2 [Caenorhabditis elegans]	71	47	252
612	2	1068	913	gi|153968	fimbriae 2 [Salmonella typhimurium]	71	55	156
613	1	1	654	gi|466778	lysine specific permease [Escherichia coli]	71	50	654
618	1	1243	623	gi|1146238	poly(A) polymerase [Bacillus subtilis]	71	52	621
630	1	1170	586	gi|1486243	unknown [Bacillus subtilis]	71	53	585
691	1	1126	641	gi|89260	comE ORF1 [Bacillus subtilis]	71	51	486
694	2	149	427	gi|12971	NADH dehydrogenase subunit V (AA 1-605) [Gallus gallus] [r|S10197|S10197 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrion (SOC1)	71	47	279
715	2	169	777	gi|1303830	YqtL [Bacillus subtilis]	71	53	609
746	2	1473	970	gi|137843	unknown [Bacillus subtilis]	71	52	504

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
748	1	1437	802	gi|1405459	YnaS [Bacillus subtilis]	71	49	636
753	1	1018	524	gi|1510389	M. jannaschii predicted coding region MJ0296 [Methanococcus jannaschii]	71	53	495
761	1	3	215	gi|475972	pentafunctional enzyme [Pneumocystis carinii]	71	47	213
783	1	1203	703	gi|536655	ORF YBR244w [Saccharomyces cerevisiae]	71	52	501
800	3	1292	987	gi|1204326	tRNA delta(2)-isopentenylpyrophosphate transferase [Haemophilus influenzae]	71	48	306
806	1	116	286	gi|1419075	cbiM gene product [Methanobacterium thermoautotrophicum]	71	50	171
931	1	973	488	gi|893358	PgsA [Bacillus subtilis]	71	56	486
1041	1	2	262	gi|1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	71	45	261
1070	1	2	172	gi|709993	hypothetical protein [Bacillus subtilis]	71	46	171
1176	1	57	365	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	71	49	309
1181	1	366	184	gi|46971	epiP gene product [Staphylococcus epidermidis]	71	50	183
1281	1	3	290	gi|153016	ORF 419 protein [Staphylococcus aureus]	71	50	288
1148	1	456	229	gi|602683	orfc [Mycoplasma capricolum]	71	48	228
2002	1	756	379	gi|1008177	ORF YJL046w [Saccharomyces cerevisiae]	71	48	378
2119	1	2	217	gi|1046088	arginyl-tRNA synthetase [Mycoplasma genitalium]	71	50	216
2418	1	3	320	gi|1499771	M. jannaschii predicted coding region MJ0936 [Methanococcus jannaschii]	71	57	318
2961	1	2	187	gi|312243	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	71	57	186
2999	2	67	306	gi|710020	nitrite reductase (nirB) [Bacillus subtilis]	71	43	240
3033	1	2	184	gi|1262315	YmaA [Bacillus subtilis]	71	57	183
3584	1	3	338	gi|401716	beta-isopropylmalate dehydrogenase [Neurospora crassa]	71	55	336
3715	2	743	399	gi|563952	gluconate permease [Bacillus licheniformis]	71	59	345
3785	1	770	387	gi|47382	acyl-CoA-dehydrogenase [Streptomyces purpurascens]	71	57	384
3875	1	541	272	gi|1001541	hypothetical protein [Synechocystis sp.]	71	38	270
4135	1	637	320	gi|142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium	71	52	318
4249	1	63	239	gi|1205363	deoxyribose aldolase [Haemophilus influenzae]	71	63	177
4508	1	530	267	gi|1119667	vitellogenin [Anolis pulchellus]	71	46	264

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
6	3	1237	2721	gi|1321788	arginine ornithine antiporter (Clostridium perfringens)	70	54	1485
11	11	6572	7486	gi|216854	P47K (Pseudomonas chlororaphis)	70	41	915
12	1	2890	1481	gi|467330	replicative DNA helicase (Bacillus subtilis)	70	49	1410
15	1	1756	893	gi|451216	Mannosephosphate isomerase (Streptococcus mutans)	70	46	864
15	2	1277	1050	gi|476092	unknown (Bacillus subtilis)	70	50	228
17	2	2132	1350	gi|145402	choline dehydrogenase (Escherichia coli)	70	52	783
21	1	2	925	gi|149516	anthranilate synthase alpha subunit (Lactococcus lactis) pir|S35124|S35124 anthranilate synthase (EC 4.1.1.27) alpha chain - actococcus lactis subsp. lactis	70	50	924
25	7	5580	6251	gi|1389549	ORF3 (Bacillus subtilis)	70	52	672
33	6	6071	7423	gi|1303875	YqhB (Bacillus subtilis)	70	51	1353
36	2	959	1594	gi|500755	methyl purine glycosylase (Mus musculus)	70	47	636
38	8	4901	5860	gi|1408507	pyrimidine nucleoside transport protein (Bacillus subtilis)	70	44	960
44	8	5312	5989	gi|1006620	hypothetical protein (Synechocystis sp.)	70	49	678
46	10	8950	10020	gi|1403126	czcD gene product (Alcaligenes eutrophus)	70	45	1071
52	2	2727	1900	gi|1486247	unknown (Bacillus subtilis)	70	53	828
52	6	4048	4656	gi|244501	esterase II-carboxylesterase (EC 3.1.1.1) (Pseudomonas fluorescens, eptide, 216 aa)	70	50	609
56	8	8460	9962	gi|1339951	small subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	70	51	1503
62	1	48	290	gi|142702	A competence protein 2 (Bacillus subtilis)	70	47	243
64	1	1080	541	gi|1204377	molybdopterin biosynthesis protein (Haemophilus influenzae)	70	47	540
70	5	5139	3595	gi|1204834	2',3'-cyclic-nucleotide 2'-phosphodiesterase (Haemophilus influenzae)	70	47	1545
91	4	7793	5466	gi|888471	methionine synthase (Catharanthus roseus)	70	56	2328
96	5	8754	7255	pir|B39096|B390	alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis	70	54	1500
110	2	767	1300	gi|145294	adenine phosphoribosyl-transferase (Escherichia coli)	70	51	534
116	6	7026	7976	gi|143607	sporulation protein (Bacillus subtilis)	70	50	951
121	8	6401	6988	gi|1107528	ctg start (Campylobacter coli)	70	45	588
131	8	6842	7936	gi|1150454	prolidase PepQ (Lactobacillus delbrueckii)	70	48	1095

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Accession	match gene name	% sim	% ident	length (nt)
135	1	2	1489	gi|311309	putative membrane-bound protein with [four times repitition of ro-Ser-Ala at the N-terminus; function unknown [Alcaligenes utrophus]	70	49	1488
128	3	418	714	gi|904181	hypothetical protein [Bacillus subtilis]	70	46	297
164	8	9344	9874	gi|49315	ORF1 gene product [Bacillus subtilis]	70	47	531
164	16	15626	16618	gi|1205212	hypothetical protein (GB:D10481_18) [Haemophilus influenzae]	70	50,	993
205	2	2735	1803	gi|1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	70	47	933
209	3	910	1386	gi|1204665	hypothetical protein (GB:X73124_26) [Haemophilus influenzae]	70	48	477
246	3	340	756	gi|215098	excisionase [Bacteriophage 154a]	70	46	417
263	7	7876	6749	gi|142540	aspartokinase II [Bacillus sp.]	70	51	1128
268	3	3212	4117	gi|340128	ORF1 [Staphylococcus aureus]	70	50	906
302	6	3201	3827	gi|147782	ruvA protein [gtg start] [Escherichia coli]	70	46	627
302	10	5879	7051	pir|C38530|C38S	queuine tRNA-ribosyltransferase (EC 2.4.2.29) - Escherichia coli	70	55	1173
313	1	2520	1414	gi|1205934	aminopeptidase a/I [Haemophilus influenzae]	70	46	1107
355	2	379	669	gi|1070013	protein-dependent [Bacillus subtilis]	70	48	291
403	1	1255	629	gi|733147	GumF [Xanthomonas campestris]	70	33	627
444	10	8770	9273	gi|1204752	high affinity ribose transport protein [Haemophilus influenzae]	70	52	504
449	1	2	1243	gi|619724	MgtE [Bacillus firmus]	70	44	1242
472	1	637	320	gi|727145	open reading frame; putative (Bacillus amyloliquefaciens) pir|B29909|B29091 hypothetical protein (bglA region) - Bacillus myoliquefaciens (fragment)	70	41	318
480	2	727	1608	gi|142560	ATP synthase gamma subunit [Bacillus megaterium]	70	44	882
524	1	2	307	gi|602292	RCH2 protein [Brassica napus]	70	45	306
525	1	823	413	gi|143372	phosphoribosyl glycinamide formyltransferase (PUR-N) [Bacillus subtilis]	70	52	411
565	4	3625	2552	gi|881434	ORF5 [Bacillus subtilis]	70	51	1074
607	4	829	1284	gi|511524	hypothetical protein (SP:P37002) [Methanococcus jannaschii]	70	50	456
633	1	1183	703	gi|431231	uracil permease [Bacillus caldolyticus]	70	53	681
646	3	1683	1309	gi|467340	unknown [Bacillus subtilis]	70	49	375
663	1	810	417	gi|1203873	Y.qz [Bacillus subtilis]	70	40	414

TABLE 2

*: aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
681	1	1488	781	gi|1001678	hypothetical protein (Synechocystis sp.)	70	53	708
708	1	2	448	sp|P33940|YOJH_	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	70	51	447
725	1	51	722	gi|1001644	hypothetical protein (Synechocystis sp.)	70	48	672
776	1	1371	787	gi|145165	putative (Escherichia coli)	70	47	585
834	1	250	783	gi|552971	NADH dehydrogenase (ndhF) (Vicia faba)	70	47'	534
865	2	1585	1379	gi|1204636	ATP-dependent helicase (Haemophilus influenzae)	70	45	207
894	1	535	269	gi|467364	DNA binding protein (probable) (Bacillus subtilis)	70	41	267
919	1	3	317	gi|1316847	CinA (Bacillus subtilis)	70	40	315
944	1	3	572	gi|709991	hypothetical protein (Bacillus subtilis)	70	44	570
988	2	772	605	gi|142441	ORF 3: putative (Bacillus subtilis)	70	50	168
1055	1	3	335	gi|529755	speC (Streptococcus pyogenes)	70	37	333
1093	1	2	904	gi|853754	ABC transporter (Bacillus subtilis)	70	49	903
1109	1	2	310	gi|1001827	hypothetical protein (Synechocystis sp.)	70	42	309
1220	1	468	235	pir|S23416|S234	epiB protein - Staphylococcus epidermidis	70	40	234
1279	1	73	348	gi|153015	FemA protein (Staphylococcus aureus)	70	47	276
1336	1	195	542	sp|P31776|PBPA_	PENICILLIN-BINDING PROTEIN 1A (PBP-1A) (PENICILLIN-BINDING PROTEIN A).	70	50	348
1537	2	232	402	gi|1146181	putative (Bacillus subtilis)	70	50	171
1574	1	451	272	gi|219630	endothelin-A receptor (Homo sapiens)	70	47	180
1640	1	690	346	gi|1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)	70	46	345
2504	1	2	286	gi|495179	transmembrane protein (Lactococcus lactis)	70	51	285
3061	1	564	301	gi|508175	EIIC domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	70	44	264
3128	1	2	199	gi|1340096	unknown (Mycobacterium tuberculosis)	70	51	198
3218	1	3	488	gi|515938	glutamate synthase (ferredoxin) (Synechocystis sp.) pir|S46957|S46957; glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	70	50	466
3323	1	794	399	gi|1154891	ATP binding protein (Phormidium laminosum)	70	52	396
1679	1	599	399	gi|529385	chromosome condensation protein (Caenorhabditis elegans)	70	30	201

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3841	1	706	398	gi|1208965	hypothetical 23.3 kd protein [Escherichia coli]	70	47	309
3929	1	3	401	gi|149435	putative [Lactococcus lactis]	70	49	399
4044	1	595	374	gi|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	70	40	222
4329	1	558	280	gi|1333951	small subunit of NADH-dependant glutamate synthase [Plectonema boryanum]	70	49	279
4422	1	576	289	gi|296464	ATPase [Lactococcus lactis]	70	57	288
4647	1	361	200	gi|166412	NADH-glutamate synthase [Medicago sativa]	70	59	162
16	8	7571	9031	gi|1499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	69	44	1461
16	9	9080	10033	gi|1353197	chloredoxin reductase [Eubacterium acidaminophilum]	69	54	954
30	1	1452	727	gi|1204910	hypothetical protein (GB:U14001_302) [Haemophilus influenzae]	69	52	726
38	4	1023	1298	gi|407773	devA gene product [Anabaena sp.]	69	41	276
44	9	5987	6595	gi|1205920	molybdate uptake system hydrophilic membrane-bound protein [Haemophilus influenzae]	69	45	609
62	15	9104	9475	gi|385178	unknown [Bacillus subtilis]	69	44	372
66	4	2402	2803	gi|1303893	YqhL [Bacillus subtilis]	69	51	402
67	15	14124	13627	gi|149647	ORF2 [Listeria monocytogenes]	69	37	498
67	17	14053	14382	gi|305002	ORF_f356 [Escherichia coli]	69	49	330
67	19	15130	15807	gi|1109684	ProV [Bacillus subtilis]	69	45	678
78	3	1447	2124	gi|1256631	putative [Bacillus subtilis]	69	53	678
78	4	4513	3725	gi|1103958	YqjO [Bacillus subtilis]	69	32	789
85	4	4521	4213	pir|E29326|E293	hypothetical protein (pur operon) - Bacillus subtilis	69	32	309
86	6	3253	2654	gi|973332	OrfC [Bacillus subtilis]	69	50	600
95	1	96	710	gi|78646A	4A11 antigen, sperm tail membrane antigen putative sucrose-specific phosphotransferase enzyme-II homolog (mice, testis, Peptide Partial, 72 aa) [Haemophilus influenzae]	69	43	615
100	7	6023	7426	gi|1205355	Na+/H+ antiporter [Haemophilus influenzae]	69	39	1404
102	2	2678	1650	gi|561690	sialoglycoprotease [Pasteurella haemolytica]	69	47	1029
103	8	12241	8537	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	69	54	3705
103	11	14987	12552	gi|710020	nitrite reductase (nirB) [Bacillus subtilis]	69	51	2436

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
112	11	8708	10168	gi|154411	hexosephosphate transport protein (Salmonella typhimurium) pir|D41853|D41853 hexose phosphate transport system protein uhpT - Salmonella typhimurium	69	51	1461
112	16	16644	17414	gi|1204435	pyruvate formate-lyase activating enzyme (Haemophilus influenzae)	69	50	771
113	2	33	953	gi|290509	o307 (Escherichia coli)	69	43	921
114	2	1537	1058	pir|A42771|A427	reticulocyte-binding protein 1 - Plasmodium vivax	69	39	480
121	6	4309	5310	gi|154633	NrdF (Bacillus subtilis)	69	53	1002
125	2	267	854	gi|413931	lpa-7d gene product (Bacillus subtilis)	69	43	588
149	27	10666	10400	pir|S28089|S280	hypothetical protein A - yeast (Zygosaccharomyces bisporus) plasmid pSB1	69	39	267
161	1	1598	813	gi|1205538	hypothetical protein (GB:U14003_302) (Haemophilus influenzae)	69	47	786
165	4	2222	4633	gi|40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) (Bacillus btilis)	69	52	2412
169	3	1210	1761	gi|296031	elongation factor Ts (Spirulina platensis)	69	45	552
175	12	8686	8339	gi|732682	FimS protein (Escherichia coli)	69	69	348
190	2	484	1671	sp|P17731|HIS8_	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE).	69	48	1188
206	1	5551	2777	gi|41750	hsdR protein (AA 1-1033) (Escherichia coli)	69	49	2775
206	4	6038	5796	gi|1256135	YbhF (Bacillus subtilis)	69	48	243
249	1	636	319	gi|1405456	YnuP (Bacillus subtilis)	69	40	318
302	8	4820	5776	gi|1001768	hypothetical protein (Synechocystis sp.)	69	48	957
324	2	7384	3893	gi|1256798	pyruvate carboxylase (Rhizobium etli)	69	53	3492
351	3	2098	1808	gi|1491664	T04H1.4 (Caenorhabditis elegans)	69	30	291
369	3	2075	2305	gi|336458	ORF (Balaenoptera acutorostrata)	69	61	231
392	3	1999	2424	gi|556015	ORF1 (Bacillus subtilis)	69	45	426
410	1	87	779	gi|155611	phosphoglyceromutase (Zymomonas mobilis)	69	58	693
421	1	2085	1129	gi|1276985	arginase (Bacillus caldovelox)	69	54	957
444	8	6713	7741	gi|1221782	purine synthesis repressor (Haemophilus influenzae)	69	40	1029
453	1	828	415	gi|1122758	unknown (Bacillus subtilis)	69	57	414
469	2	3286	2246	gi|1458228	mutY homolog (Homo sapiens)	69	44	1041

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
509	3	1730	1371	gi|49224	URF 4 [Synechococcus sp.]	69	39	360
520	5	3023	2823	gi|726427	similar to D. melanogaster HST101-2 protein (PIR:S34154) Caenorhabditis elegans	69	39	201
531	1	26	760	gi|509672	repressor protein (bacteriophage Tuc2009)	69	33	235
589	1	107	253	gi|169101	17.9 kDa heat shock protein (hsp17.9) [Pisum sativum]	69	52	147
594	2	597	1391	gi|142783	DNA photolyase [Bacillus firmus]	69	48	795
604	4	2476	2114	gi|413930	ipa-6d gene product [Bacillus subtilis]	69	45	363
607	1	2	313	gi|1236103	W08D2.3 [Caenorhabditis elegans]	69	47	312
607	2	590	312	gi|536715	ORF YBR275c [Saccharomyces cerevisiae]	69	39	279
734	1	864	433	gi|467327	unknown [Bacillus subtilis]	69	44	432
759	1	3	338	gi|1009367	Respiratory nitrate reductase [Bacillus subtilis]	69	50	336
761	2	392	586	gi|3508	Leucyl-tRNA synthetase (cytoplasmic) [Saccharomyces cerevisiae] |1370140 ORF YPL160w [Saccharomyces cerevisiae]	69	46	195
802	1	72	1013	gi|143064	ferrochelatase [Bacillus subtilis]	69	55	942
816	1	2573	1368	gi|1510268	restriction modification system S subunit [Methanococcus jannaschii]	69	45	1206
838	2	133	387	gi|1255371	coded for by C. elegans cDNA yk34a9.5; coded for by C. elegans cDNA yk34a9.1; similar to guanylate kinase [Caenorhabditis elegans]	69	46	255
851	2	745	1005	gi|288998	scA gene product [Antithamnion sp.]	69	39	261
867	1	535	269	gi|1070014	protein-dependent [Bacillus subtilis]	69	47	267
995	1	954	478	gi|1205569	transcription elongation factor [Haemophilus influenzae]	69	53	477
999	1	1009	506	gi|899254	predicted trithorax protein [Drosophila virilis]	69	21	504
1127	1	1315	659	gi|1205434	H. influenzae predicted coding region HI1191 [Haemophilus influenzae]	69	56	657
1138	1	248	460	gi|1510646	M. jannaschii predicted coding region MJ0568 [Methanococcus jannaschii]	69	48	213
2928	1	3	401	gi|290503	glutamate permease [Escherichia coli]	69	41	399
3090	1	444	223	gi|1204987	DNA polymerase III, alpha chain [Haemophilus influenzae]	69	36	222
3817	1	2	400	gi|1483199	peptide-synthetase [Amycolatopsis mediterranei]	69	45	399
3833	1	667	335	gi|1524193	unknown [Mycobacterium tuberculosis]	69	46	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4079	1	747	400	gi|546918	orfY 3' of comK (Bacillus subtilis, E26, Peptide Partial, 140 aa) pir|S43612|S43612 hypothetical protein Y - Bacillus subtilis sp|P40398|YHXD_BACSU HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFY FRAGMENT).	69	64	348
4115	2	215	400	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	69	59	186
4139	1	1	333	gi|1208451	hypothetical protein (Synechocystis sp.)	69	36	333
4258	1	457	230	gi|496158	restriction-modification enzyme subunit M1 (Mycoplasma pulmonis) pir|S49395|S49395 HsdM1 protein - Mycoplasma pulmonis (SGC3)	69	43	228
4317	1	90	374	gi|413967	lpa-43d gene product (Bacillus subtilis)	69	44	285
4465	1	3	293	gi|396296	similar to phosphotransferase system enzyme II (Escherichia coli) sp|P32672|PTWC_ECOLI PTS SYSTEM, FRUCTOSE-LIKE-2 IIC COMPONENT PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).	69	49	291
3	1	2102	1193	gi|1109865	ProW (Bacillus subtilis)	68	46	1110
15	4	2592	2074	gi|807973	unknown (Saccharomyces cerevisiae)	68	45	519
31	8	6328	8772	gi|290642	ATPase (Enterococcus hirae)	68	48	2445
40	2	1115	750	gi|606342	ORF_0622) reading frame open far upstream of start; possible rameshift, linking to previous ORF (Escherichia coli)	68	55	366
46	9	6886	8415	gi|155276	aldehyde dehydrogenase (Vibrio cholerae)	68	44	1530
48	3	3643	3404	gi|285608	241k polyprotein (Apple stem grooving virus)	68	47	240
48	4	3536	4132	gi|1045937	M. genitalium predicted coding region MG246 (Mycoplasma genitalium)	68	39	597
53	10	11671	10685	gi|1303952	YqjA (Bacillus subtilis)	68	46	987
70	9	7346	8155	gi|147198	phnR protein (Escherichia coli)	68	40	810
89	4	1899	2966	gi|145173	35 kDa protein (Escherichia coli)	68	43	1068
108	1	2187	1150	gi|38722	precursor (aa -20 to 391) (Acinetobacter calcoaceticus) tr|A29277|A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus	68	57	1038
112	5	2666	3622	gi|153724	MalC (Streptococcus pneumoniae)	68	55	957
116	7	7865	8638	gi|143608	sporulation protein (Bacillus subtilis)	68	48	774
118	3	2484	3698	gi|1103805	YqeR (Bacillus subtilis)	68	46	1215
120	2	1424	1594	sp|P18038|CYSJ_	SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA-COMPONENT (EC 1.8.1.2) (SIR-FP).	68	45	171
129	1	1	1011	gi|396307	argininosuccinate lyase (Escherichia coli)	68	50	1011

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
112	3	1867	2739	gi|216267	ORF2 [Bacillus megaterium]	68	48	873
134	2	848	1012	gi|147545	DNA recombinase [Escherichia coli]	68	50	165
141	2	372	614	gi|872116	atl (stress inducible protein) [Glycine max]	68	36	243
149	7	2454	2260	gi|145774	hsp70 protein (dnak gene) [Escherichia coli]	68	48	195
155	2	1776	1534	gi|216583	ORF1 [Escherichia coli]	68	38	243
158	3	1826	3289	sp|P33940|YOJH_	HYPOTHETICAL 54.3 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	68	51	1464
169	6	2749	3318	gi|1401402	unknown [Mycobacterium tuberculosis]	68	46	570
175	10	9158	7365	gi|1072395	phaA gene product [Rhizobium meliloti]	68	51	1794
188	7	4184	5434	gi|1173863	3-ketoacyl-ACP synthase II [Vibrio harveyi]	68	48	1251
189	3	907	1665	gi|467383	DNA binding protein (probable) [Bacillus subtilis]	68	55	759
206	5	7683	6709	gi|1256138	YbbT [Bacillus subtilis]	68	48	975
206	8	10425	12176	gi|452687	pyruvate decarboxylase [Saccharomyces cerevisiae]	68	48	1752
212	8	3421	3648	gi|1369941	c1 gene product [Bacteriophage B1]	68	39	228
214	8	5457	6482	gi|1420467	ORF YOR196c [Saccharomyces cerevisiae]	68	45	1026
237	4	2507	3088	gi|149381	HisH [Lactococcus lactis]	68	46	582
243	5	5540	4542	gi|1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	68	47	999
262	1	3	164	gi|150974	4-oxalocrotonate tautomerase [Pseudomonas putida]	68	42	162
262	2	1984	1118	gi|147744	PSR [Enterococcus hirae]	68	49	867
276	6	3702	3139	sp|P30750|ABC_E	ATP-BINDING PROTEIN ABC (FRAGMENT).	68	50	564
106	6	6345	5725	gi|1256617	adenine phosphoribosyltransferase [Bacillus subtilis]	68	53	621
333	3	4599	3850	gi|467473	unknown [Bacillus subtilis]	68	45	750
365	6	5017	4838	gi|1130643	F22B3.3 [Caenorhabditis elegans]	68	45	180
376	2	549	1646	gi|1277026	DAPA aminotransferase [Bacillus subtilis]	68	51	1098
405	1	1741	872	gi|1303917	YqiB [Bacillus subtilis]	68	47	870
406	2	853	539	gi|1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	68	44	315
426	6	3558	3391	gi|624632	GltL [Escherichia coli]	68	48	168
438	1	108	329	gi|146923	nitrogenase reductase [Escherichia coli]	68	43	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
443	1	476	240	gi|535810	hipuricase [Campylobacter jejuni]	68	42	237
443	2	518	1015	gi|1204742	H. influenzae predicted coding region HI0491 [Haemophilus influenzae]	68	48	498
443	5	4447	3779	gi|609660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir|S49455|S49455 deoxyribose-phosphate aldolase [EC 4.1.2.4] - acilius subtilis	68	55	669
476	2	200	1184	gi|571345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp|P45860|YMIE_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA NTERGENIC REGION.	68	45	945
486	2	1876	1046	gi|147328	transport protein [Escherichia coli]	68	41	831
517	3	1764	2084	gi|522809	orf2 [Bacteriophage A2]	68	64	321
572	1	2	571	sp|P19217|YO5L_	HYPOTHETICAL 6.8 KD PROTEIN IN NRDC-PK INTERGENIC REGION.	68	47	570
646	1	914	459	gi|413982	ipa-58r gene product [Bacillus subtilis]	68	52	456
659	3	1668	1901	gi|1407541	C13D9.8 [Caenorhabditis elegans]	68	36	234
864	5	1510	1716	gi|145774	hsp70 protein (dnaK gene) [Escherichia coli]	68	48	207
920	1	860	432	gi|1510416	hypothetical protein (SP:P31466) [Methanococcus jannaschii]	68	54	429
952	1	1096	611	gi|C23456	reductase [Leishmania major]	68	46	486
970	1	91	402	gi|1354775	pfoS/R [Treponema pallidum]	68	46	312
1028	1	1064	534	gi|410117	[diaminopimelate decarboxylase [Bacillus subtilis]	68	47	531
1029	1	428	216	gi|1135714	Plasmodium falciparum mRNA for asparagine-rich antigen (clone 17C1) [Plasmodium falciparum]	68	31	213
1058	1	692	348	gi|81649	epiC gene product [Staphylococcus epidermidis]	68	46	345
1096	2	665	465	gi|143434	Rho Factor [Bacillus subtilis]	68	43	201
1308	1	2	694	gi|69939	group B oligopeptidase.PepB [Streptococcus agalactiae]	68	50	693
1679	1	2	238	gi|17205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	68	53	237
2039	1	3	383	gi|153898	transport protein [Salmonella typhimurium]	68	51	381
2077	1	3	126	pir|C33496|C334	hisC homolog - Bacillus subtilis	68	47	324
2112	1	613	374	gi|64884	lamin LII [Xenopus laevis]	68	50	240
2273	1	793	398	gi|581648	epiB gene product [Staphylococcus epidermidis]	68	45	196
2948	1	2	385	gi|216869	branched-chain amino acid transport carrier [Pseudomonas aeruginosa] pir|A38534|A38534 branched-chain amino acid transport protein braZ Pseudomonas aeruginosa	68	41	384

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match Acession	match gene name	% sim	% ident	length (nt)
2955	1	768	400	gi|904179	hypothetical protein [Bacillus subtilis]	68	49	369
2981	1	572	288	gi|508979	GTP-binding protein [Bacillus subtilis]	68	48	285
3014	1	584	294	gi|1524394	ORF-2 upstream of gbaAB operon [Bacillus subtilis]	68	45	291
3042	1	336	169	gi|1204696	fructose-permease IIBC component [Haemophilus influenzae]	68	53	168
3108	1	103	258	gi|217855	heat-shock protein [Arabidopsis thaliana]	68	48	156
3639	1	919	461	gi|1510490	nitrate transport permease protein [Methanococcus jannaschii]	68	47	459
3657	1	1	330	gi|155369	PTS enzyme-II fructose [Xanthomonas campestris]	68	48	330
3823	1	780	391	gi|603768	Hutl protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi|603768 Hutl protein, imidazolone-5-propionate hydrolase Bacillus subtilis]	68	54	390
3982	1	2	277	gi|149435	putative [Lactococcus lactis]	68	47	276
4051	1	1	342	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	48	342
4089	1	12	209	gi|1353678	heavy-metal transporting P-type ATPase [Proteus mirabilis]	68	47	198
4143	1	47	187	gi|603769	HutU protein, urocanase [Bacillus subtilis]	68	55	141
4148	1	2	352	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	51	351
4173	1	2	382	gi|1041097	Pyruvate kinase [Bacillus psychrophilus]	68	48	381
4182	1	498	250	gi|413968	ipa-44d gene product [Bacillus subtilis]	68	50	249
4362	2	148	318	gi|450688	hsdM gene of EcoprrI gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	44	171
5	11	9493	8300	gi|143727	putative [Bacillus subtilis]	67	46	1194
31	11	10318	9833	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	67	41	486
32	3	1560	3155	gi|1098557	renal sodium/dicarboxylate cotransporter [Homo sapiens]	67	46	1596
32	5	4945	4145	gi|1510720	prephenate dehydratase [Methanococcus jannaschii]	67	51	801
36	5	5350	4268	gi|146216	45% identity with the product of the ORF6 gene from the Erwinia herbicola carotenoid biosynthesis cluster; putative [Bacillus subtilis]	67	58	1083
44	7	4492	5304	gi|1006621	hypothetical protein [Synechocystis sp.]	67	43	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	7	3943	8481	gi|304131	glutamate synthase large subunit precursor (Azospirillum brasilense) pir|B46602|B46602 glutamate synthase (NADPH) (EC 1.4.1.13) alpha chain - Azospirillum brasilense	67	52	4539
56	12	13923	14678	gi|1000453	TreR (Bacillus subtilis)	67	48	756
62	8	5092	4757	gi|1113949	orf3 (Bacillus. C-125, alkali-sensitive mutant 18224, Peptide Mutant, 112 aa)	67	45	336
62	10	7570	6338	gi|154655	Na/H antiporter system (Bacillus alcalophilus)	67	49	1233
99	3	2119	3321	gi|1204349	hypothetical protein (GB:GB:D90212_3) (Haemophilus influenzae)	67	50	1203
102	9	5695	7176	gi|149432	putative (Lactococcus lactis)	67	51	1482
103	11	14549	14049	gi|1408497	L29D gene product (Bacillus subtilis)	67	48	501
109	15	14821	13982	gi|413976	ipa-52r gene product (Bacillus subtilis)	67	49	840
109	17	14811	15194	gi|413983	ipa-59d gene product (Bacillus subtilis)	67	29	384
121	4	1713	2153	gi|1262335	YmaA (Bacillus subtilis)	67	54	441
122	1	1	1149	gi|143047	ORFB (Bacillus subtilis)	67	35	1149
124	5	4060	3518	gi|556885	Unknown (Bacillus subtilis)	67	47	543
131	2	4584	3589	gi|1046081	hypothetical protein (GB:D26185_10) (Mycoplasma genitalium)	67	30	996
140	3	2899	2297	gi|146549	kdpC (Escherichia coli)	67	45	601
142	4	5409	4198	gi|1212775	GTP cyclohydrolase II (Bacillus amyloliquefaciens)	67	55	1212
147	5	2913	2374	gi|1303709	YrkJ (Bacillus subtilis)	67	44	540
152	8	6341	6673	gi|1137841	unknown (Bacillus subtilis)	67	48	333
161	4	2720	3763	gi|496319	SphX (Synechococcus sp.)	67	47	1044
163	6	1989	3428	gi|595681	2-oxoglutarate/malate translocator (Spinacia oleracea)	67	47	1440
193	3	1351	1626	gi|1511101	shikimate 5-dehydrogenase (Methanococcus jannaschii)	67	53	276
200	2	917	2179	gi|142439	ATP-dependent nuclease (Bacillus subtilis)	67	48	1263
206	10	12445	12801	sp|P37347|YECD_	HYPOTHETICAL 21.8 KD PROTEIN IN ASPS 5'REGION.	67	47	357
206	11	13047	14432	gi|732813	branched-chain amino acid carrier (Lactobacillus delbrueckii)	67	46	1386
208	2	1321	809	gi|1033037	100 kDa heat shock protein (Hsp100) (Leishmania major)	67	36	513
238	3	1039	2052	gi|809542	CbrB protein (Erwinia chrysanthemi)	67	42	1014

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
246	2	176	367	gi|215098	excisionase (Bacteriophage 1S6a)	67	37	192
276	2	2260	1412	gi|303560	ORF271 (Escherichia coli)	67	50	849
297	6	2223	3056	gi|142784	CtaA protein (Bacillus firmus)	67	46	834
307	7	5220	4186	gi|1070013	protein-dependent (Bacillus subtilis)	67	43	1035
316	1	16	1028	gi|1161061	dioxygenase (Methylobacterium extorquens)	67	57	993
324	3	5650	5030	gi|1469784	putative cell division protein ftsW [Enterococcus hirae]	67	49	621
336	1	524	264	gi|171122	urea amidolyase [Saccharomyces cerevisiae]	67	45	261
340	1	108	1194	(sp)|P30053|SYH_S	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIGASE) (HISRS).	67	47	1287
364	3	4890	3592	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	46	1299
365	1	2940	2113	gi|1296823	orf2 gene product [Lactobacillus helveticus]	67	47	828
367	2	325	918	gi|1039479	ORFU [Lactococcus lactis]	67	47	594
395	3	666	1271	gi|1204516	hypothetical protein (GB:U00014_4) [Haemophilus influenzae]	67	55	606
415	1	1800	901	gi|882579	CG Site No. 29739 [Escherichia coli]	67	46	900
419	1	1799	903	gi|520752	putative [Bacillus subtilis]	67	48	897
474	1	2	796	pir|A46906	argininosuccinate synthetase (Streptomyces clavuligerus) pir|S57659|S57659 argininosuccinate synthase (EC 6.3.4.5) - streptomyces clavuligerus	67	49	795
485	2	1921	2226	gi|144434	Rho factor [Bacillus subtilis]	67	43	306
596	1	1728	865	gi|1303853	YqgF [Bacillus subtilis]	67	47	864
700	1	433	218	gi|1204628	hypothetical protein (SP:P21498) [Haemophilus influenzae]	67	47	216
806	2	249	647	gi|677947	AppC [Bacillus subtilis]	67	51	399
828	2	340	900	gi|777761	lrrA [Synechococcus sp.]	67	37	561
833	1	1407	916	gi|142996	regulatory protein [Bacillus subtilis]	67	41	492
856	1	1555	779	gi|780224	2K970.2 [Caenorhabditis elegans]	67	38	777
888	1	1614	850	gi|437315	TTG start codon [Bacillus licheniformis]	67	40	765
1034	1	1190	597	gi|1205113	hypothetical protein (GB:L19201_19) [Haemophilus influenzae]	67	45	594
1062	1	636	319	gi|1303850	YqgC [Bacillus subtilis]	67	41	318
1067	1	918	460	pir|AJ2950|A229	probable reductase protein - Leishmania major	67	54	459

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1358	1	3	293	gi|1001369	hypothetical protein (Synechocystis sp.)	67	44	291
2181	1	3	302	gi|1510416	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	67	48	300
3000	1	1	507	gi|517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogenes)	67	56	507
3066	1	464	234	gi|308861	GTG start codon (Lactococcus lactis)	67	46	231
3087	1	454	251	gi|1205366	oligopeptide transport ATP-binding protein (Haemophilus influenzae)	67	44	204
3101	1	2	256	gi|1531541	uroporphyrinogen III methyltransferase (Zea mays)	67	55	255
3598	1	728	393	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	56	336
3765	2	584	366	gi|557489	menD (Bacillus subtilis)	67	45	219
3788	1	658	398	pir|S52915|S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	67	45	261
3883	1	2	265	gi|704397	cystathionine beta-lyase (Arabidopsis thaliana)	67	46	264
3926	1	2	340	gi|1463199	peptide-synthetase (Amycolatopsis mediterranei)	67	44	339
4417	1	82	396	gi|1205337	ribonucleotide transport ATP-binding protein (Haemophilus influenzae)	67	46	315
2	3	3075	3989	gi|535348	CodV (Bacillus subtilis)	66	42	915
15	6	2273	2542	gi|46491	SmtB (Synechococcus PCC7942)	66	37	270
11	9	8059	7826	gi|1292046	mucin (Homo sapiens)	66	44	234
31	10	9034	9258	gi|1204545	mercury scavenger protein (Haemophilus influenzae)	66	40	225
32	6	6347	5253	gi|998342	inducible nitric oxide synthase (Gallus gallus)	66	47	1095
44	13	8856	10124	gi|1510751	molybdenum cofactor biosynthesis moeA protein (Methanococcus jannaschii)	66	46	1269
48	2	1276	2868	gi|150209	ORF 1 (Mycoplasma mycoides)	66	40	1593
59	8	7178	8428	gi|665999	hypothetical protein (Bacillus subtilis)	66	47	1251
62	7	5143	4370	gi|1072398	phaD gene product (Rhizobium meliloti)	66	40	774
70	14	11693	10998	gi|809660	deoxyribose-phosphate aldolase (Bacillus subtilis) pir|S49455|S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	66	55	696
76	1	1	1305	gi|142440	ATP-dependent nuclease (Bacillus subtilis)	66	42	1305
91	6	9236	8205	gi|704397	cystathionine beta-lyase (Arabidopsis thaliana)	66	43	1032
102	5	3810	3265	gi|1204323	hypothetical protein (SP:P31805) (Haemophilus influenzae)	66	41	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
103	4	3418	2732	gi|971144	nitrate reductase gamma subunit [Bacillus subtilis] sp|P42177|NARI_BACSU NITRATE REDUCTASE GAMMA CHAIN (EC 1.7.99.4). gi|1009369 Respiratory nitrate reductase [Bacillus subtilis] (SUB -160)	66	48	687
109	6	4243	4674	gi|170886	glucosamine-6-phosphate deaminase [Candida albicans] pir|A46652|A46652 glucosamine-6-phosphate isomerase (EC 5.3.1.10) - east (Candida albicans)	66	45	432
112	17	17491	17712	gi|1323179	ORF YGR111w [Saccharomyces cerevisiae]	66	33	222
116	2	4667	2637	gi|1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	66	43	2031
150	5	3189	2989	gi|1146224	putative [Bacillus subtilis]	66	30	201
172	5	3264	3662	gi|1755152	highly hydrophobic integral membrane protein [Bacillus subtilis] sp|P42953|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN ACG.	66	41	399
174	5	4592	3723	gi|1146241	pantothenate synthetase [Bacillus subtilis]	66	49	870
175	4	3209	2880	gi|642655	unknown [Rhizobium meliloti]	66	29	330
175	11	8743	7994	gi|854655	Na/H antiporter system [Bacillus alcalophilus]	66	43	750
190	5	7079	5727	gi|451072	di-tripeptide transporter [Lactococcus lactis]	66	40	1353
195	15	13919	13713	gi|1322411	unknown [Mycobacterium tuberculosis]	66	42	207
217	13	2822	2595	gi|1143542	alternative stop codon [Rattus norvegicus]	66	36	228
233	9	7133	6135	gi|1458327	F08F3.4 gene product [Caenorhabditis elegans]	66	47	999
238	1	41	1041	gi|809541	CbrA protein [Erwinia chrysanthemi]	66	42	999
241	1	2102	1053	gi|153067	peptidoglycan hydrolase [Staphylococcus aureus]	66	53	1050
261	1	1178	648	gi|1510859	M. jannaschii predicted coding region MJ0790 [Methanococcus jannaschii]	66	40	531
263	3	3731	2973	gi|1205865	tetrahydrodipicolinate N-succinyltransferase [Haemophilus influenzae]	66	47	759
272	8	6548	5484	gi|882101	high affinity nickel transporter [Alcaligenes eutrophus] sp|P23516|HOXN_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN.	66	44	1065
276	3	2805	2104	gi|1208965	hypothetical 23.3 kd protein [Escherichia coli]	66	47	702
278	2	2830	1784	gi|1488662	phosphatase-associated protein [Bacillus subtilis]	65	48	1047
278	3	3830	2952	gi|303560	ORF271 [Escherichia coli]	66	45	879
279	2	3894	2218	gi|1185289	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1- carboxylate synthase [Bacillus subtilis]	66	48	1677
288	4	2535	2275	gi|1256625	putative [Bacillus subtilis]	65	42	261
292	2	1133	942	gi|1511604	M. jannaschii predicted coding region MJ1651 [Methanococcus jannaschii]	66	30	192

TABLE 2

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (aa)
294	1	1116	559	gi|216314	esterase [Bacillus stearothermophilus]	66	45	558
297	4	2913	1978	gi|994794	cytochrome a assembly facto [Bacillus subtilis] sp|P24009|COXX_BACSU PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.	66	45	936
316	4	2051	2882	gi|1107839	alginate lyase [Pseudomonas aeruginosa]	66	40	610
338	4	2460	2302	gi|520750	biotin synthetase [Bacillus sphaericus]	66	55	159
339	1	1214	785	gi|467468	?, 8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Bacillus subtilis]	66	52	480
363	1	3	863	gi|581649	epiC gene product [Staphylococcus epidermidis]	66	47	661
366	2	232	483	gi|1103505	unknown [Schizosaccharomyces pombe]	66	53	252
367	4	2468	1845	sp|P20692|TYRA_	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).	66	50	624
372	3	2150	1599	gi|467416	unknown [Bacillus subtilis]	66	38	552
378	1	212	1009	gi|147309	purine nucleoside phosphorylase [Escherichia coli]	66	50	798
401	1	1	462	gi|388263	p-aminobenzoic acid synthase [Streptomyces griseus] pir|JN0531|JN0531 p-aminobenzoic acid synthase - Streptomyces riseus	66	46	462
404	7	4826	5234	gi|606744	cytidine deaminase [Bacillus subtilis]	66	51	429
411	2	1738	1103	gi|1460081	unknown [Mycobacterium tuberculosis]	66	44	636
420	1	2	541	gi|1046024	Na+ ATPase subunit J [Mycoplasma genitalium]	66	49	540
431	1	1	858	gi|1500008	M. jannaschii predicted coding region MJ1154 [Methanococcus jannaschii]	66	50	858
443	7	5679	5299	gi|852076	hrgA [Bacillus subtilis]	66	46	381
444	3	3405	2413	gi|151047	lysostaphin (ttg start codon) [Staphylococcus simulans] pir|A25881|A25881 lysostaphin precursor - Staphylococcus simulans sp|P10547|LSTP_STASI LYSOSTAPHIN PRECURSOR (EC 3.5.1.-).	66	51	993
561	1	956	480	gi|1204905	DNA-3-methyladenine glycosidase I [Haemophilus influenzae]	66	45	477
562	3	1066	1383	gi|1046082	M. genitalium predicted coding region MG372 [Mycoplasma genitalium]	66	52	318
576	1	11	724	gi|305014	ORF_o234 [Escherichia coli]	66	43	714
577	3	1190	903	gi|1001353	hypothetical protein [Synechocystis sp.]	66	52	288
584	1	2	331	sp|P24204|YEBA_	HYPOTHETICAL 46.7 KD PROTEIN IN MSBB-RUVB INTERGENIC REGION (ORFU).	66	48	330
592	1	1410	706	gi|928839	ORF266; putative [Lactococcus lactis phage BK5-T]	66	51	705
601	1	1433	720	gi|1488655	novel antigen; orf-2 [Staphylococcus aureus]	66	55	714

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
619	3	468	845	gi|746571	similar to M. musculus transport system membrane protein, Nramp PIR:A40739) and S. cerevisiae SNF1 protein (PIR:A45154) Caenorhabditis elegans)	66	45	378
706	2	561	355	gi|804808	unknown protein (Rattus norvegicus)	66	46	207
734	2	673	512	gi|1519085	phosphatidylcholine binding immunoglobulin heavy chain IgM variable region (Mus musculus)	66	60	162
740	1	3	317	gi|1209272	argininosuccinate:lyase (Campylobacter jejuni)	66	47	315
764	1	310	747	gi|435296	alkaline phosphatase like protein (Lactococcus lactis) pir|S39339|S39339 alkaline phosphatase-like protein - Lactococcus actis	66	42	438
852	1	338	171	gi|536955	CG Site No. 361 (Escherichia coli)	66	43	168
886	1	3	158	gi|289272	[ferrichrome-binding protein (Bacillus subtilis)	66	44	156
889	1	462	232	gi|813061	HCMVUL77 (AA 1-642) (Human cytomegalovirus)	66	66	231
893	1	2	247	gi|149008	putative (Helicobacter pylori)	66	45	246
900	1	1425	733	gi|580842	[F] (Bacillus subtilis)	66	51	693
906	2	2300	1473	gi|790945	aryl-alcohol dehydrogenase (Bacillus subtilis)	66	53	828
947	1	79	549	gi|410117	diaminopimelate decarboxylase (Bacillus subtilis)	66	47	471
950	1	1100	552	gi|48713	orf145 (Staphylococcus aureus)	66	35	549
955	2	89	475	gi|1204390	uridine kinase (uridine monophosphokinase) (Haemophilus influenzae)	66	50	387
981	2	1308	997	gi|457146	rhoptry protein (Plasmodium yoelii)	66	38	312
986	1	25	315	gi|305002	ORF_f356 (Escherichia coli)	66	31	291
1057	1	3	203	gi|1303853	YrgP (Bacillus subtilis)	66	40	201
1087	1	1	294	gi|575913	unknown (Saccharomyces cerevisiae)	66	53	294
1105	1	1	231	gi|1045799	methylgalactoside permease ATP-binding protein (Mycoplasma genitalium)	66	46	231
1128	1	2	574	gi|1001493	hypothetical protein (Synechocystis sp.)	66	46	573
1150	1	498	250	gi|1499034	M. jannaschii predicted coding region MJ0255 (Methanococcus jannaschii)	66	40	249
1180	2	707	453	gi|215908	DNA polymerase (gp43) (Bacteriophage T4)	66	46	255
1208	1	1123	587	gi|1256653	DNA-binding protein (Bacillus subtilis)	66	58	537
1342	1	1	402	gi|1208474	hypothetical protein (Synechocystis sp.)	66	53	402
1761	2	589	398	gi|215811	tail fiber protein (Bacteriophage T3)	66	50	192

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1983	1	499	251	gi|1045935	DNA helicase II (Mycoplasma genitalium)	66	40	249
2103	2	176	400	gi|929798	precursor for the major merozoite surface antigens (Plasmodium alciparum)	66	46	225
2341	1	373	188	gi|1256623	exodeoxyribonuclease (Bacillus subtilis)	66	38	186
2458	1	325	164	gi|1019410	unknown (Schizosaccharomyces pombe)	66	47	162
2505	1	468	235	gi|1510394	putative transcriptional regulator (Methanococcus jannaschii)	66	39	234
2525	1	558	280	gi|1000695	cytotoxin L (Clostridium sordellii)	66	44	279
2935	1	3	275	gi|765073	autolysin (Staphylococcus aureus)	66	47	273
3005	1	114	305	gi|1205784	heterocyst maturation protein (Haemophilus influenzae)	66	46	192
3068	1	80	277	gi|1303813	YqaW (Bacillus subtilis)	66	42	198
3071	1	1	189	gi|1070014	protein-dependent (Bacillus subtilis)	66	41	189
3081	1	404	225	gi|984212	unknown (Schizosaccharomyces pombe)	66	44	180
3090	2	580	386	gi|1204987	DNA polymerase III, alpha chain (Haemophilus influenzae)	66	48	195
3118	1	1	387	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	66	49	387
3739	1	798	400	gi|1109684	ProV (Bacillus subtilis)	66	47	399
3796	1	402	202	gi|853760	acyl-CoA dehydrogenase (Bacillus subtilis)	66	60	201
3924	1	121	347	gi|563952	gluconate permease (Bacillus licheniformis)	66	46	249
4240	1	1	350	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	66	51	348
4604	1	7	234	pir|A26713|BHHC	hemocyanin subunit II - Atlantic horseshoe crab	66	46	228
4	9	8845	9750	gi|145646	cynR (Escherichia coli)	65	35	906
6	5	2708	3565	gi|887824	ORF_o310 (Escherichia coli)	65	47	858
13	1	1993	998	gi|143402	recombination protein (ttg start codon) (Bacillus subtilis) gi|1303923 RecN (Bacillus subtilis)	65	44	996
15	7	2493	3524	gi|1403126	czcD gene product (Alcaligenes eutrophus)	65	38	1032
18	3	1908	1372	gi|349187	acyltransferase (Saccharomyces cerevisiae)	65	50	537
21	3	1467	2492	gi|149518	phosphoribosyl anthranilate transferase (Lactococcus lactis) pir|S35126|S35126 anthranilate phosphoribosyltransferase (BC .4.2.18) - Lactococcus lactis subsp. lactis	65	52	1026
25	4	3374	4312	gi|502420	malonyl-CoA:Acyl carrier protein transacylase (Bacillus subtilis)	65	44	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
27	2	390	626	gi|1212729	YqbJ [Bacillus subtilis]	65	45	237
31	12	11040	10187	gi|509245	D-hydroxyisocaproate dehydrogenase [Lactobacillus delbrueckii]	65	41	654
38	24	19172	19528	gi|547:	M-protein [Flaveria trinervia]	65	41	357
44	2	790	1746	gi|405882	yeiK [Escherichia coli]	65	46	957
44	12	9356	8832	gi|1205905	molybdenum cofactor biosynthesis protein [Haemophilus influenzae]	65	50	525
45	8	6635	7588	gi|493074	AppA protein [Salmonella typhimurium]	65	46	954
51	2	580	1503	gi|580897	OppB gene product [Bacillus subtilis]	65	45	924
52	1	225	953	gi|1205518	NAD(P)H-flavin oxidoreductase [Haemophilus influenzae]	65	45	729
55	4	1339	1058	pir|A44459|A444	troponin T beta Tnf-5 - rabbit	65	41	282
67	9	7421	8272	gi|143607	sporulation protein [Bacillus subtilis]	65	42	852
73	5	4446	5375	gi|1204896	lysophospholipase L2 [Haemophilus influenzae]	65	37	930
74	1	954	478	gi|1204844	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	65	50	477
77	1	2	757	gi|1046082	M. genitalium predicted coding region MG372 [Mycoplasma genitalium]	65	46	756
77	2	795	1433	gi|1222116	permease [Haemophilus influenzae]	65	37	639
81	3	4728	3454	gi|1001708	hypothetical protein [Synechocystis sp.]	65	49	1275
91	7	8548	8357	gi|1339263	cystathionine beta-lyase [Emericella nidulans]	65	40	192
98	3	1608	1988	gi|467423	unknown [Bacillus subtilis]	65	36	381
98	4	2250	2987	gi|467424	unknown [Bacillus subtilis]	65	45	738
102	3	2598	2119	gi|1511532	N-terminal acetyltransferase complex, subunit ARD1 [Methanococcus jannaschii]	65	39	480
102	4	3647	2862	gi|1204637	H. influenzae predicted coding region HI0388 [Haemophilus influenzae]	65	32	786
103	9	10851	9841	gi|142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium	65	47	1011
103	10	10439	10119	gi|710021	nitrite reductase (nirD) [Bacillus subtilis]	65	51	321
106	2	262	1140	gi|39881	ORF 311 (AA 1-311) [Bacillus subtilis]	65	44	879
109	5	3909	4268	gi|1204399	glucosamine-6-phosphate deaminase protein [Haemophilus influenzae]	65	44	360
109	10	7165	8595	gi|536955	CG Site No. 361 [Escherichia coli]	65	41	1431

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
110	4	3688	3915	gi|407881	stringent response-like protein [Streptococcus equisimilis] pir|S39975|S39975 stringent response-like protein - Streptococcus quisimilis	65	45	228
110	5	3882	4295	gi|407880	ORF1 [Streptococcus equisimilis]	65	50	414
110	6	4231	4380	gi|1135574	orf2 [Streptomyces griseus]	65	56	150
112	10	9218	8640	gi|1204571	H. influenzae predicted coding region HI0318 [Haemophilus influenzae]	65	52	579
112	12	12049	11288	gi|710496	[transcriptional activator protein [Bacillus brevis]	65	32	762
125	1	2	202	gi|1151158	repeat organellar protein [Plasmodium chabaudi]	65	39	201
126	1	3	422	gi|37589	precursor [Homo sapiens]	65	46	420
127	11	10733	12658	gi|1064809	homologous to splHTRA_ECOLI [Bacillus subtilis]	65	41	1926
143	8	7543	7004	gi|216513	mutator mutT (AT-GC transversion) [Escherichia coli]	65	56	540
145	5	3587	3838	gi|1209768	D02_orf569 [Mycoplasma pneumoniae]	65	27	252
150	4	3482	2841	gi|1146225	putative [Bacillus subtilis]	65	37	642
166	1	3858	1948	gi|148304	beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir|A42296|A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790)	65	50	1911
188	6	3195	4178	gi|351943	ORF3; putative [Rhodobacter capsulatus]	65	46	984
189	9	4982	4785	gi|15RR12	ORF IV (AA 1-489) [Pigment mosaic virus]	65	40	198
195	6	7900	5272	gi|145220	alanyl-tRNA synthetase [Escherichia coli]	65	49	2637
195	7	10599	8104	gi|882711	exonuclease V alpha-subunit [Escherichia coli]	65	38	2496
206	16	16896	18191	gi|408115	ornithine acetyltransferase [Bacillus subtilis]	65	53	1296
217	4	3844	3215	gi|1205974	5'guanylate kinase [Haemophilus influenzae]	65	41	630
220	4	5265	3751	gi|580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13464|TAGR_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPIA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TEC1101C ACID BIOSYNTHESIS PROTEIN E).	65	40	1515
236	5	2327	3709	gi|1146200	DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis]	65	46	1383
237	3	1902	2513	gi|149379	HisBd [Lactococcus lactis]	65	46	612
241	4	4968	4195	gi|1205308	ribonuclease HII (EC 3.1.26.4) (RNASE HII) [Haemophilus influenzae]	65	50	774
252	1	1278	940	gi|1204989	hypothetical protein (GB:U00022_9) [Haemophilus influenzae]	65	40	339
261	5	4780	3794	gi|145927	facD [Escherichia coli]	65	43	987

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
274	1	3	278	gi|496558	ortX (Bacillus subtilis)	65	42	276
301	2	982	815	gi|67418	unknown (Bacillus subtilis)	65	45	168
307	4	3586	2864	gi|1070014	protein-dependent (Bacillus subtilis)	65	40	723
315	2	2286	1399	gi|146911	N-acetylglucosamine transport protein (Escherichia coli) pir|B29895|WQEC2N phosphotransferase system enzyme II (EC .7.1.69), N-acetylglucosamine-specific - Escherichia coli sp|P09323|PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC OMPONENT (EIIA	65	50	888
338	5	4120	3170	gi|1277029	biotin synthase (Bacillus subtilis)	65	49	951
343	3	1490	2800	gi|143264	membrane-associated protein (Bacillus subtilis)	65	48	1311
344	4	2761	2531	gi|1050540	tRNA-glutamine synthetase (Lupinus luteus)	65	34	231
358	3	3421	3621	gi|146220	NAD+ dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)	65	47	201
364	1	238	699	gi|1340128	ORF1 (Staphylococcus aureus)	65	51	462
379	1	1	576	gi|143331	alkaline phosphatase regulatory protein (Bacillus subtilis) pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-).	65	40	576
379	3	3666	4346	gi|143268	dihydrolipoamide transsuccinylase (odhB; EC 2.3.1.61) (Bacillus ubtilis)	65	50	681
428	1	187	483	gi|1420465	ORF YOR195v (Saccharomyces cerevisiae)	65	45	297
438	2	272	838	gi|143498	dcgS protein (Bacillus subtilis)	65	34	567
444	11	9280	10215	gi|1204756	ribokinase (Haemophilus influenzae)	65	47	936
449	2	1241	1531	gi|599848	Na/H antiporter homolog (Lactococcus lactis)	65	41	291
478	2	1452	865	gi|1045942	glycyl-tRNA synthetase (Mycoplasma genitalium)	65	39	588
479	1	1032	517	gi|1498192	putative (Pseudomonas aeruginosa)	65	40	516
480	6	4312	5637	gi|415662	UDP-N-acetylglucosamine 1-carboxyvinyl transferase (Acinetobacter alcoaceticus)	65	48	1326
484	1	2	430	gi|146551	transmembrane protein (KdpD) (Escherichia coli)	65	44	429
499	1	54	932	gi|603456	reductase (Leishmania major)	65	53	879
505	1	914	459	gi|1518853	OafA (Salmonella typhimurium)	65	39	456
571	2	1509	883	gi|149399	open reading frame upstream glnE (Escherichia coli) ir|S37754|S37754 hypothetical protein XE (glnE 5' region) - cherichia coli	65	44	627
611	2	506	270	gi|10961	RAP-2 (Plasmodium falciparum)	65	40	237

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
705	1	564	283	gi|710020	nitrite reductase (nirB) [Bacillus subtilis]	65	52	282
712	1	1	177	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	177
712	2	196	354	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	159
743	1	2	631	gi|310631	ATP binding protein [Streptococcus gordonii]	65	45	630
749	2	393	779	gi|167374	single strand DNA binding protein [Bacillus subtilis]	65	29	387
762	1	1698	850	gi|160399	multidrug resistance protein [Plasmodium falciparum]	65	48	849
788	1	85	315	gi|1129096	unknown protein [Bacillus sp.]	65	35	231
850	1	1	408	gi|1006604	hypothetical protein [Synechocystis sp.]	65	37	408
908	1	1	444	gi|1119546	2362 [Saccharomyces cerevisiae]	65	46	444
925	1	1	174	gi|1256653	DNA-binding protein [Bacillus subtilis]	65	54	174
1031	1	26	232	gi|238657	AppC acytochrome d oxidase, subunit I homolog [Escherichia coli, K12, peptide, 514 aa]	65	47	207
1037	1	414	262	gi|1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	65	46	153
1053	1	388	175	gi|642655	unknown [Rhizobium meliloti]	65	34	174
1149	1	1399	752	gi|1162980	ribulose-5-phosphate 3-epimerase [Spinacia oleracea]	65	48	648
1214	1	881	495	gi|1205959	lactam utilization protein [Haemophilus influenzae]	65	45	387
1276	1	476	276	pir|S35492|S354	site-specific DNA-methyltransferase StsI (EC 2.1.1.-) - Streptococcus sanguis	65	35	201
1276	2	900	577	gi|473794	'ORF' [Escherichia coli]	65	34	324
2057	1	272	138	gi|633699	TreH [Yersinia enterocolitica]	65	21	135
2521	1	336	169	gi|1045789	hypothetical protein (GB:U14003_76) [Mycoplasma genitalium]	65	41	168
2974	1	590	297	gi|152052	enantiomerase-selective amidase [Rhodococcus sp.]	65	45	294
3031	1	106	154	pir|JQ1024|JQ10	hypothetical 10K protein (DmRP140 5' region) - fruit fly (Drosophila melanogaster)	65	47	153
3069	1	3	278	gi|44906	product homologous to E.coli thioredoxin reductase; J.Biol.Chem. 1988] 261:9015-9019, and to F52a protein of alkyl hydroperoxide reductase from S.typhimurium; J.Biol.Chem. (1990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum]	65	66	276
3146	1	282	142	gi|49315	[ORF] gene product [Bacillus subtilis]	65	47	141
3170	1	679	341	gi|1507711	indolepyruvate decarboxylase [Erwinia herbicola]	65	44	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3546	1	1	303	u|450688	hsdM gene of Ecopri gene product [Escherichia coli] pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	65	42	303
3782	1	2	128	gi|166412	NADH-glutamate synthase [Medicago sativa]	65	42	327
3990	1	174	189	gi|1009366	Respiratory nitrate reductase [Bacillus subtilis]	65	53	186
4032	1	613	308	gi|1323127	ORF YGR087c [Saccharomyces cerevisiae]	65	50	306
4278	2	726	364	gi|1197667	vitellogenin [Anolis pulchellus]	65	42	363
19	4	4259	5518	gi|145727	deaD [Escherichia coli]	64	45	1260
19	6	7639	6926	gi|1016232	ycf27 gene product [Cyanophora paradoxa]	64	36	714
20	8	7053	6454	gi|765073	autolysin [Staphylococcus aureus]	64	47	600
31	13	12706	11537	gi|414009	ipa-85d gene product [Bacillus subtilis]	64	45	1170
33	4	2388	4364	gi|1204696	fructose-permease IIBC component [Haemophilus influenzae]	64	47	1977
36	3	1871	3013	gi|290503	glutamate permease [Escherichia coli]	64	40	1143
37	6	4065	4409	gi|39815	orf 2 gene product [Bacillus subtilis]	64	46	345
45	9	7852	8760	gi|1230585	nucleotide sugar epimerase [Vibrio cholerae O139]	64	53	909
53	3	1540	1899	gi|1303961	YqjJ [Bacillus subtilis]	64	50	360
56	6	4793	3855	gi|457514	gltC [Bacillus subtilis]	64	45	939
56	24	30002	30247	gi|470331	similar to zinc fingers [Caenorhabditis elegans]	64	42	246
62	4	2759	2421	gi|642655	unknown [Rhizobium meliloti]	64	28	339
85	6	7178	6027	gi|457702	5-aminoimidazole ribonucleotide-carboxilase [Pichia methanolica] pir|S39112|S39112 phosphoribosylaminoimidazole carboxylase (EC .1.1.21) - yeast (Pichia methanolica)	64	46	1152
96	9	9251	10030	gi|1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	64	42	780
100	1	1	600	gi|765073	autolysin [Staphylococcus aureus]	64	44	600
106	5	3868	4854	gi|466778	lysine specific permease [Escherichia coli]	64	46	987
123	2	838	554	gi|467484	unknown [Bacillus subtilis]	64	47	285
127	8	7514	7810	gi|210061	serotype-specific antigen [African horse sickness virus] pir|S27891|S27891 capsid protein VP2 - African horse sickness virus	64	28	297
131	7	7134	6721	gi|1511160	M. jannaschii predicted coding region MJ1163 [Methanococcus jannaschii]	64	46	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
142	5	5455	4817	gi|1173517	riboflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]	64	44	639
143	1	709	356	pir|A32950|A329	probable reductase protein - Leishmania major	64	52	354
149	10	3555	3295	gi|398151	major surface antigen MSG2 [Pneumocystis carinii]	64	44	261
154	4	3134	2307	gi|988587	DinP [Escherichia coli]	64	50	828
161	5	3855	4880	gi|901304	ORF72 [Bacillus subtilis]	64	37	1026
165	1	33	791	gi|467483	unknown [Bacillus subtilis]	64	38	759
175	6	6355	4714	gi|1072398	phaD gene product [Rhizobium meliloti]	64	42	1512
188	3	2042	2500	gi|1001961	MHC class II analog [Staphylococcus aureus]	64	45	459
195	14	13667	13446	gi|396380	No definition line found [Escherichia coli]	64	47	222
206	15	16429	16938	gi|304134	argC [Bacillus stearothermophilus]	64	49	510
215	1	560	282	gi|142359	ORF 6 [Azotobacter vinelandii]	64	39	279
243	7	7818	6928	gi|141014	ipa-90d gene product [Bacillus subtilis]	64	49	891
258	2	1330	845	gi|664754	P17 [Listeria monocytogenes]	64	38	486
259	1	462	232	gi|1149663	M. Jannaschii predicted coding region MJ0837 [Methanococcus jannaschii]	64	52	231
263	6	6565	5567	gi|142828	aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp|Q04797|DHAS_BACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC .2.1.11) (ASA DEHYDROGENASE)	64	48	999
271	1	3	1163	gi|467091	hflX; B2215_C2_202 [Mycobacterium leprae]	64	44	1161
280	1	173	1450	gi|1303839	YqfR [Bacillus subtilis]	64	43	1278
293	1	2552	1267	gi|147345	primosomal protein n' [Escherichia coli]	64	45	1266
295	2	742	1488	gi|459266	Potential membrane spanning protein [Staphylococcus hominis] pir|S42932|S42932 potential membrane spanning protein - staphylococcus hominis	64	39	747
301	5	1625	1446	gi|580835	lysine decarboxylase [Bacillus subtilis]	64	35	180
315	4	5064	3949	gi|143396	quinol oxidase [Bacillus subtilis]	64	45	1116
321	1	1264	635	gi|710496	transcriptional activator protein [Bacillus brevis]	64	41	630
333	5	4520	4239	gi|1314295	ORF2: putative 19 kDa protein [Listeria monocytogenes]	64	43	282
342	1	1	549	gi|142940	ftsA [Bacillus subtilis]	64	38	549
353	3	2878	2324	gi|537049	ORF_0470 [Escherichia coli]	64	44	555

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
379	2	827	3658	pir|S25295|A328	oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) - Bacillus subtilis	64	47	2832
404	6	4429	4839	pir|A36933|A369	diacylglycerol kinase homolog - Streptococcus mutans	64	35	411
407	1	2020	1113	gi|969026	orfX (Bacillus subtilis)	64	41	888
425	1	1109	591	gi|1146177	phosphotransferase system glucose-specific enzyme II (Bacillus subtilis)	64	44	519
443	6	4082	4798	gi|147309	purine nucleoside phosphorylase (Escherichia coli)	64	51	717
450	2	1035	1604	gi|c06176	ORF_o162 (Escherichia coli)	64	38	570
470	5	1680	6107	gi|1369948	host interacting protein (Bacteriophage B1)	64	45	4428
486	4	1911	1471	gi|1205582	spermidine/putrescine transport system permease protein (Haemophilus influenzae)	64	35	441
497	1	2217	1159	sp|P36929|PMU_E	PMU PROTEIN.	64	38	1059
501	1	3	410	gi|142450	ahrC protein (Bacillus subtilis)	64	38	408
514	1	3	290	gi|1204496	H. influenzae predicted coding region HI0238 (Haemophilus influenzae)	64	34	288
551	4	3162	3323	gi|1204511	bacterioferritin comigratory protein (Haemophilus influenzae)	64	41	162
603	4	759	956	gi|755823	HADH dehydrogenase F (Streptogyna americana)	64	35	198
653	2	940	746	gi|211234	dicarboxylic amino acids Dip5p permease (Saccharomyces cerevisiae)	64	41	195
660	3	3H01	2257	sp|P46133|YDAIL	HYPOTHETICAL PROTEIN IN OGT 5'REGION (FRAGMENT).	64	39	1545
695	1	11	502	gi|1001383	hypothetical protein (Synechocystis sp.)	64	41	492
702	1	3	752	gi|142865	DNA primase (Bacillus subtilis)	64	46	750
826	1	1	339	gi|971336	arginyl tRNA synthetase (Bacillus subtilis)	64	50	339
838	1	1831	917	gi|154775	pfoS/R (Treponema pallidum)	64	41	915
864	3	675	944	gi|39833	cyclomaltodextrin glucanotransferase (Bacillus stearothermophilus) 1|39835 cyclomaltodextrin glucanotransferase (Bacillus earothermophilus)	64	47	270
887	1	3	677	gi|153002	enterotoxin type E precursor (Staphylococcus aureus) pir|A28179|A28179 enterotoxin E precursor - Staphylococcus aureus sp|P12993|ETXE_STAAU ENTEROTOXIN TYPE E PRECURSOR (SEE).	64	46	675
928	2	1172	963	gi|311976	fibrinogen-binding protein (Staphylococcus aureus) pir|S34270|S34270 fibrinogen-binding protein - Staphylococcus ureus	64	41	210
1049	2	800	606	gi|1049115	Rap60 (Bacillus subtilis)	64	42	195
1067	2	999	748	gi|1151072	RhdA precursor (Haemophilus ducreyl)	64	50	252

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1120	1	50	202	gi|142439	ATP-dependent nuclease [Bacillus subtilis]	64	30	153
1125	1	751	377	gi|581648	epiB gene product [Staphylococcus epidermidis]	64	44	375
1688	1	402	214	pir|A01365|TVMS	transforming protein K-ras - mouse	64	47	189
2472	1	2	358	gi|487282	Na+ -ATPase subunit J [Enterococcus hirae]	64	36	357
2989	1	520	356	gi|304134	argC [Bacillus stearothermophilus]	64	50	165
3013	1	630	352	gi|551699	cytochrome oxidase subunit I [Bacillus firmus]	64	51	279
3034	1	546	274	gi|204349	hypothetical protein (GB:GB:D90212_3) [Haemophilus influenzae]	64	50	273
3197	1	613	308	gi|1009356	Respiratory nitrate reductase [Bacillus subtilis]	64	46	306
1303	1	90	362	gi|1107839	alginate lyase [Pseudomonas aeruginosa]	64	43	273
3852	2	82	288	gi|216746	D-lactate dehydrogenase [Lactobacillus plantarum]	64	42	207
3868	1	1	312	gi|149435	putative [Lactococcus lactis]	64	48	312
3918	1	660	331	gi|5532	acetyl-CoA acyltransferase [Yarrowia lipolytica]	64	46	330
4000	1	112	378	gi|994688	unknown [Saccharomyces cerevisiae]	64	44	267
4009	1	81	368	gi|39372	grsB gene product [Bacillus brevis]	64	41	288
4166	1	2	349	gi|149435	putative [Lactococcus lactis]	64	46	348
4366	1	2	107	gi|216267	ORF2 [Bacillus megaterium]	64	44	306
4457	1	2	400	gi|197667	vitellogenin [Anolis pulchellus]	64	43	399
11	3	1539	2438	gi|438228	ORF C [Staphylococcus aureus]	63	32	900
24	7	5611	5423	gi|1369943	al gene product [Bacteriophage Bl]	63	34	189
29	1	1	390	gi|467441	expressed at the end of exponential growth under condtions in which he enzymes of the TCA cycle are repressed [Bacillus subtilis] gi|467441 expressed at the end of exponential growth under ondtions in which the enzymes of the TCA cycle are repressed Bacil	63	43	390
31	6	6329	5712	gi|496943	ORF [Saccharomyces cerevisiae]	63	47	618
44	23	14669	15019	pir|A04446|QQEC	hypothetical protein F-92 - Escherichia coli	63	36	351
48	6	4403	6250	gi|43498	pyruvate synthase [Halobacterium halobium]	63	42	1848
50	5	3869	4738	gi|411967	ipa-43d gene product [Bacillus subtilis]	63	43	870
53	6	6764	5742	gi|474176	regulator protein [Staphylococcus xylosus]	63	49	1023

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	14	15880	17607	gi|467409	DNA polymerase III subunit (Bacillus subtilis)	63	44	1728
57	11	7945	7376	gi|537036	ORF_o158 (Escherichia coli)	63	39	570
62	3	2479	2114	gi|142656	unknown (Rhizobium meliloti)	63	41	366
70	8	6562	7353	gi|1399821	PhoC (Rhizobium meliloti)	63	46	792
75	2	223	927	gi|149376	HisQ (Lactococcus lactis)	63	45	705
78	5	4912	4403	gi|411950	ipa-26d gene product (Bacillus subtilis)	63	42	510
91	5	9076	7220	gi|466997	metH2; B2126_C1_157 (Mycobacterium leprae)	63	41	1857
91	8	10566	9448	gi|1204344	cystathionine gamma-synthase (Haemophilus influenzae)	63	45	1119
120	1	21	1508	gi|882657	sulfite reductase (NADPH) flavoprotein beta subunit (Escherichia coli)	63	46	1488
120	4	2722	4125	gi|665994	hypothetical protein (Bacillus subtilis)	63	34	1404
127	7	6064	7566	gi|40162	murE gene product (Bacillus subtilis)	63	44	1503
149	6	2321	2106	gi|148503	dnaK (Erysipelothrix rhusiopathiae)	63	40	216
149	26	10445	10170	gi|4870	ORF 2, has similarity to DNA polymerase [Saccharomyces kluyveri] r[S15961|S15961 hypothetical protein 2 - yeast (Saccharomyces yveri) plasmid pSKL	63	42	276
164	2	507	1298	gi|145476	CDP-diglyceride synthetase (Escherichia coli)	63	44	792
166	6	9909	8164	gi|151912	fructose enzyme II (Rhodobacter capsulatus)	63	41	1746
169	4	1704	1886	gi|152886	elongation factor Ts (tsf) (Spiroplasma citri)	63	48	183
188	5	3145	2951	gi|1334547	GIY COI i14 grp IB protein (Podospora anserina)	63	42	195
195	13	11767	12804	gi|606100	ORF_o335 (Escherichia coli)	63	40	1038
201	2	607	2283	gi|433534	arginyl-tRNA synthetase [Corynebacterium glutamicum] pir|A49936|A49936 arginine--tRNA ligase (EC 6.1.1.19) - orynebacterium glutamicum	63	46	1677
206	14	15893	16689	gi|580828	N-acetyl-glutamate-gamma-semialdehyde dehydrogenase (Bacillus subtilis)	63	49	597
220	5	7769	5766	gi|216334	secA protein (Bacillus subtilis)	63	42	2004
221	1	74	907	gi|677945	AppA (Bacillus subtilis)	63	42	834
227	3	944	1708	gi|1510558	cobyric acid synthase (Methanococcus jannaschii)	63	46	765
261	2	804	1070	gi|486511	ORF YKR054c (Saccharomyces cerevisiae)	63	45	267
269	2	3606	1960	gi|148221	DNA-dependent ATPase, DNA helicase [Escherichia coli] pir|JS0137|BVECRQ recQ protein - Escherichia coli	63	42	1647

TABLE 2

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
278	8	7417	6176	gi|699273	cystathionine gamma-synthase [Mycobacterium leprae] sp|P46807|METB_MYCLE CYSTATHIONINE GAMMA-SYNTHASE (EC 4.2.99.9) O-SUCCINYLHOMOSERINE (THIOL)-LYASE].	63	41	1242
287	2	738	1733	gi|405133	putative [Bacillus subtilis]			996
295	1	2	748	gi|1239983	hypothetical protein [Bacillus subtilis]	63	38	747
328	3	2148	3134	gi|45302	carrier protein (AA 1 - 437) [Pseudomonas aeruginosa] tr|S11497|S11497 branched-chain amino acid tr.sport protein braB - eudomonas aeruginosa	63	41	987
362	2	1626	1216	sp|P35136|SERA_	D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).	63	36	411
404	1	326	1051	gi|1303816	YqeZ [Bacillus subtilis]	63	38	726
405	1	2101	1715	gi|1303914	YqhV [Bacillus subtilis]	63	35	387
406	1	451	227	gi|142152	sulfate permease (gtg start codon) [Synechococcus PCC6301] pir|A30301|GRYCS7 sulfate transport protein - Synechococcus sp. PCC 7942)	63	42	225
415	2	1048	2718	gi|1205402	transport ATP-binding protein [Haemophilus influenzae]	63	43	1671
426	4	3575	2679	gi|393268	29-kiloDalton protein [Streptococcus pneumoniae] sp|P42362|P29K_STRPN 29 KD MEMBRANE PROTEIN IN PSAA 5'REGION ORF).	63	41	897
505	3	1347	2195	gi|1418999	orf4 [Lactobacillus sake]	63	39	849
507	1	2	574	gi|546917	comX [Bacillus subtilis, E26, Peptide, 192 aa]	63	40	573
562	2	146	1084	gi|43985	nifS-like gene [Lactobacillus delbrueckii]	63	35	939
675	1	427	215	gi|1510994	serine aminotransferase [Methanococcus jannaschii]	63	45	213
686	1	3	230	gi|517356	nitrate reductase (NADH) [Lotus japonicus]	63	29	228
701	1	3	392	gi|881940	NorQ protein [Paracoccus denitrificans]	63	52	390
720	1	2	400	gi|47168	open reading frame [Streptomyces lividans]	63	41	399
779	1	571	287	gi|1261912	unknown [Mycobacterium tuberculosis]	63	35	285
907	1	22	321	gi|149445	ORF1 [Lactococcus lactis]	63	41	300
972	1	794	399	gi|1511235	M. jannaschii predicted coding region MJ1232 [Methanococcus jannaschii]	63	27	396
1065	1	1154	618	gi|1204277	hypothetical protein (GB:U00019_14) [Haemophilus influenzae]	63	27	537
1094	1	3	542	gi|790943	urea amidolyase [Bacillus subtilis]	63	38	540
1108	1	482		pir|S49892|S498	regulation protein - Bacillus subtilis	63	39	440
1113	1	1231	617	gi|493017	endocarditis specific antigen [Enterococcus faecalis]	63	44	615

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1300	1	3	695	sp|P33940|YOJH_	HYPOTHETICAL 54.1 KD PROTEIN IN ECO-ALKB INTERGENIC REGION.	63	46	693
1325	1	1	204	gi|928989	p100 protein [Borrelia burgdorferi]	63	30	204
1814	1	3	245	gi|1301914	YqhY [Bacillus subtilis]	63	34	243
2021	1	498	250	pir|C33496|C334	hisC homolog - Bacillus subtilis	63	46	249
2325	1	2	193	gi|436132	product is similar to TnpA of transposon Tn554 from Staphylococcus aureus [Clostridium butyricum]	63	40	192
2335	1	1	195	gi|1184298	flagellar MS-ring protein [Borrelia burgdorferi]	63	47	195
2406	1	451	227	gi|1041785	rhopty protein [Plasmodium yoelii]	63	33	225
2961	2	136	360	gi|312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	63	52	225
2965	1	1	402	gi|1407784	orf-1; novel antigen [Staphylococcus aureus]	63	50	402
2987	1	583	293	gi|1224069	amidase [Moraxella catarrhalis]	63	25	291
2994	1	266	135	gi|836646	phosphoribosylformaino-praic ketoisomerase [Rhodobacter phaeroides]	63	51	132
3043	1	440	252	gi|1180237	phenylacetaldehyde dehydrogenase [Escherichia coli]	63	40	189
3078	1	609	400	gi|1487982	intrinsic membrane protein [Mycoplasma hominis]	63	36	210
3139	1	2	217	gi|439126	glutamate synthase (NADPH) [Azospirillum brasilense] pir|A49916|A49916 glutamate synthase (NADPH) (EC 1.4.1.13) - zospirillum brasilense	63	47	216
1625	1	793	198	gi|623073	ORF160; putative [bacteriophage L5-II]	63	48	496
3658	1	1	399	gi|1303697	YrkA [Bacillus subtilis]	63	37	399
3659	1	3	395	gi|1256135	YbbF [Bacillus subtilis]	63	48	393
3783	1	720	361	gi|1256902	Pyruvate decarboxylase (isozyme 2 (Swiss Prot. accession number P16467) [Saccharomyces cerevisiae]	63	34	360
3900	1	338	171	sp|P10537|AMYB_	BETA-AMYLASE (EC 3.2.1.2) (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE).	63	54	168
4309	1	3	176	pir|A37967|A379	neural cell adhesion molecule Ng-CAM precursor - chicken	63	57	174
4367	1	1	195	gi|1121932	Per6p gene product [Pichia pastoris]	63	30	195
4432	1	1	312	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC.1.1.1.88) Pseudomonas sp.	63	51	312
4468	1	6	308	gi|296464	ATPase [Lactococcus lactis]	63	36	303
33	3	1411	2400	gi|153675	tagatose 6-P kinase [Streptococcus mutans]	62	44	990
36	9	5985	6218	gi|1490521	hHSH3 [Homo sapiens]	62	51	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
37	1	2	721	gi|1107531	ceuE gene product (Campylobacter coli)	62	33	720
38	15	110912	111589	gi|1222058	H. influenzae predicted coding region HIN1279 (Haemophilus influenzae)	62	38	678
38	25	119526	120329	gi|695280	ORF2 (Alcaligenes eutrophus)	62	41	804
57	2	2523	1780	gi|171234	orf1 (Haemophilus influenzae)	62	55	744
57	9	6646	6350	gi|508174	EIIB domain of PTS-dependent Gat transport and phosphorylation Escherichia coli	62	35	297
58	1	2	559	gi|755152	highly hydrophobic integral membrane protein (Bacillus subtilis) sp|P42953|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	62	34	558
67	10	8250	9014	gi|4706RJ	Shows similarity with ATP-binding proteins from other ABC-transport parons. Swiss Prot Accession Numbers P24137, V08007, P04285, P24116 Escherichia coli	62	33	765
69	8	8315	7494	gi|46816	actVA 4 gene product (Streptomyces coelicolor)	62	44	822
80	3	1793	1320	gi|39993	UDP-N-acetylmuramoylalanine--D-glutamate ligase (Bacillus subtilis)	62	43	474
87	7	7034	9205	gi|217191	5'-nucleotidase precursor (Vibrio parahaemolyticus)	62	48	2172
100	3	4051	3089	gi|1511047	phosphoglycerate dehydrogenase (Methanococcus jannaschii)	62	42	963
102	1	2	520	gi|153655	mismatch repair protein (Streptococcus pneumoniae) pir|C28667|C28667 DNA mismatch repair protein hexA - Streptococcus neumoniae	62	34	519
112	2	466	1068	gi|153741	ATP-binding protein (Streptococcus mutans)	63	37	603
114	7	6855	7562	gi|1204866	L-fucose operon activator (Haemophilus influenzae)	62	38	708
116	4	6823	5633	gi|677947	AppC (Bacillus subtilis)	62	37	1191
124	8	6855	6004	gi|853777	product similar to E.coli PRFA2 protein (Bacillus subtilis) pir|S55438|S55438 ywAB protein - Bacillus subtilis sp|P45873|HEMX_BACSU POSSIBLE PROTOPORPHYRINOGEN OXIDASE (EC .1.3.-.).	62	44	852
148	1	24	554	gi|467456	unknown (Bacillus subtilis)	62	30	531
149	20	7591	6725	gi|1205807	replicative DNA helicase (Haemophilus influenzae)	62	41	867
163	3	1503	1153	gi|40067	x gene product (Bacillus sphaericus)	62	42	351
164	15	14673	15632	gi|42219	P35 gene product (AA 1 - 314) (Escherichia coli)	62	38	960
165	2	1166	1447	gi|403936	phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified cloning vector	62	38	282
166	2	2084	5089	gi|1308861	CTG start codon (Lactococcus lactis)	62	44	1006
171	1	1225	614	gi|1046053	hypothetical protein (SP:P32049) (Mycoplasma genitalium)	62	41	612

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
183	1	2521	1310	gi|143045	hemY (Bacillus subtilis)	62	45	1212
200	1	3	956	gi|142439	ATP-dependent nuclease (Bacillus subtilis)	62	32	954
237	2	935	1966	gi|41695	hisC protein (Escherichia coli)	62	44	1032
261	3	4008	2605	gi|143121	ORF A; putative (Bacillus firmus)	62	42	1404
299	8	4477	4719	gi|467441	expressed at the end of exponential growth under conditions in which he enzymes of the TCA cycle are repressed (Bacillus subtilis) gi|467441 expressed at the end of exponential growth under onditions in which the enzymes of the TCA cycle are repressed Bacil	62	47	243
304	6	5018	3819	gi|153015	FemA protein (Staphylococcus aureus)	62	43	1200
324	1	2	262	gi|142717	cytochrome aa3 controlling protein (Bacillus subtilis) pir|A33960|A33960 cta protein - Bacillus subtilis sp|P12946|CTAA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN.	62	30	261
325	2	269	1207	gi|581088	methionyl-tRNA formyltransferase (Escherichia coli)	62	39	939
332	6	4494	4631	gi|149960	uridine 5'-monophosphate synthase (Methanococcus jannaschii)	62	36	264
355	1	2	370	gi|45925	fecB (Escherichia coli)	62	32	369
365	8	6628	6804	gi|41943	lipa-19d gene product (Bacillus subtilis)	62	54	177
369	2	2744	1626	pir|A43577|A435	regulatory protein pfoR - Clostridium perfringens	62	42	1119
370	1	34	264	gi|40065	beta-glucosidase (Clostridium thermocellum)	62	37	231
415	3	2709	3176	gi|1205401	transport ATP-binding protein (Haemophilus influenzae)	62	35	468
429	1	1578	790	gi|1046024	Na+ ATPase subunit J (Mycoplasma genitalium)	62	40	789
444	2	704	1369	gi|581510	nodulation gene; integral membrane protein; homology to Rhizobium equinlosarum nodi (Rhizobium loti)	62	37	666
477	2	751	1869	pir|A18440|A484	ring-infected erythrocyte surface antigen 2, RESA-2 - Plasmodium falciparum	62	44	1119
485	1	241	1707	gi|17934	betaine aldehyd dehydrogenase (Beta vulgaris)	62	43	1467
487	3	1141	1311	gi|149445	ORF1 (Lactococcus lactis)	62	31	171
494	2	1134	1313	gi|166835	ribulose bisphosphate carboxylase/oxygenase activase (Arabidopsis haliana)	62	37	180
518	1	193	882	gi|153491	O-methyltransferase (Streptomyces glaucescens)	62	39	690
534	2	369	2522	gi|148D429	putative transcriptional regulator (Bacillus stearothermophilus)	62	35	2154
551	6	4371	4820	gi|511113	ferric uptake regulation protein (Campylobacter jejuni)	62	37	450
574	1	1	570	gi|153000	enterotoxin B (Staphylococcus aureus)	62	43	570

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
590	2	344	1171	gi|40367	ORFC (Clostridium acetobutylicum)	62	37	828
655	1	396	810	gi|147195	phnB protein (Escherichia coli)	62	44	435
656	1	2	478	gi|1205451	cell division inhibitor (Haemophilus influenzae)	62	36	477
676	1	692	348	gi|1511613	methyl coenzyme M reductase system, component A2 (Methanococcus jannaschii)	62	36	345
687	1	493	248	gi|49272	Asparaginase (Bacillus licheniformis)	62	46	246
700	2	267	944	gi|1205822	hypothetical protein (GB:X75627_4) (Haemophilus influenzae)	62	40	678
840	2	1715	1041	gi|1045865	M. genitalium predicted coding region MG181 (Mycoplasma genitalium)	62	36	675
864	4	898	1491	gi|1114332	deoxyuridine nucleotidohydrolase (Homo sapiens)	62	38	594
916	1	35	400	gi|413931	ipa-7d gene product (Bacillus subtilis)	62	45	366
1071	1	1	771	gi|1510649	aspartokinase I (Methanococcus jannaschii)	62	40	771
1084	1	19	609	gi|688011	AgX-1 antigen [human, infertile patient, testis, Peptide, 505 aa]	62	39	591
1103	1	3	203	gi|581261	ORF homologous to E.coli motB [Herpetosiphon aurantiacus] pir|S14030|S14030 Hypothetical protein - Herpetosiphon aurantiacus (fragment)	62	51	201
1217	1	463	233	gi|460025	ORF2, putative (Streptococcus pneumoniae)	62	41	231
1533	1	644	414	gi|413968	ipa-44d gene product (Bacillus subtilis)	62	48	231
1537	1	3	257	gi|1510641	alanyl-tRNA synthetase (Methanococcus jannaschii)	62	29	255
2267	1	1	161	gi|485956	mrpC gene product (Proteus mirabilis)	62	45	159
2286	1	3	245	gi|285708	nontoxic component (Clostridium botulinum)	62	31	243
2484	1	331	167	gi|142092	DNA-repair protein (recA) (Anabaena variabilis)	62	35	165
2490	1	798	400	gi|581648	epiB gene product (Staphylococcus epidermidis)	62	42	399
3016	1	596	300	gi|710022	uroporphyrinogen III (Bacillus subtilis)	62	51	297
3116	1	1	213	gi|466883	nifS; B1496_C2_193 (Mycobacterium leprae)	62	44	213
3297	1	823	413	gi|475715	acetyl coenzyme A acetyltransferase (thiolase) (Clostridium acetobutylicum)	62	42	411
3609	1	31	276	gi|1408502	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus (Bacillus subtilis)	62	48	246
3665	2	584	402	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	62	40	183
3733	1	3	374	gi|1353197	thioredoxin reductase (Eubacterium acidaminophilum)	62	42	372

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3898	1	1	237	gi|153675	tagatose 6-P kinase [Streptococcus mutans]	62	45	237
4027	1	283	143	gi|330705	homologue to gene 30 (as 1-59); putative [Bovine herpesvirus 4]	62	43	141
4109	1	727	365	gi|41748	hsdM protein (AA 1-520) [Escherichia coli]	62	45	363
4303	1	1	303	gi|1301813	YqeW [Bacillus subtilis]	62	43	303
4380	1	530	267	gi|1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	62	55	264
4494	1	2	256	gi|510692	enterotoxin H [Staphylococcus aureus]	62	34	255
4598	1	411	223	gi|763513	ORF4; putative [Streptomyces violaceoruber]	62	45	189
4624	1	1	222	gi|41748	hsdM protein (AA 1-520) [Escherichia coli]	62	45	222
5	5	4288	3912	gi|928831	ORF9S; putative [Lactococcus lactis phage BK5-T]	61	36	357
11	1	320	162	pir|C33356|C33	prothymosin alpha homolog (clone 32) - human (fragment)	61	33	159
16	11	10091	11938	gi|1205391	hypothetical protein (SP:P33995) [Haemophilus influenzae]	61	44	948
22	5	281	801	gi|1066504	exo-beta 1,3 glucanase [Cochliobolus carbonum]	61	50	519
18	3	616	1107	gi|1510864	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	61	41	492
45	4	3082	4038	gi|1109686	ProX [Bacillus subtilis]	61	45	957
48	8	7118	7504	gi|498839	ORF2 [Clostridium perfringens]	61	33	387
51	9	4605	5570	gi|388269	traC [Plasmid pAD1]	61	42	966
60	6	1689	2243	gi|1205893	hypothetical protein (GB:U00011_3) [Haemophilus influenzae]	61	32	555
62	9	5559	5122	gi|854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	61	38	438
67	5	4330	5646	gi|466612	nikA [Escherichia coli]	61	36	1317
74	2	2400	1504	gi|1204846	carbamate kinase [Haemophilus influenzae]	61	40	897
85	1	2198	1101	gi|1498756	amidophosphoribosyltransferase PurF [Rhizobium etli]	61	41	1098
86	4	1995	1582	gi|1499931	M. jannaschii predicted coding region MJ1083 [Methanococcus jannaschii]	61	44	414
97	1	74	649	gi|1518679	orf [Bacillus subtilis]	61	44	576
99	2	2454	1990	gi|413958	ipa-34d gene product [Bacillus subtilis]	61	18	465
124	7	6223	5123	gi|556881	Similar to Saccharomyces cerevisiae SUA5 protein [Bacillus subtilis] pir|S49358|S49358 ipc-29d protein - Bacillus subtilis sp|P39153|YRLC_BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SPOIIR-GLVC MYERGENIC REGION.	61	46	1101
125	4	1668	2531	gi|1491643	ORFA gene product [Chloroflexus aurantiacus]	61	43	864

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
132	1	1250	627	pir|PQ0259|PQ02	hypothetical protein I - (Enterococcus faecalis plasmid pAM-beta-1 (fragment)	61	43	624
149	9	3617	3075	gi|1144332	deoxyuridine nucleotidohydrolase [Homo sapiens]	61	40	543
149	22	8690	7869	gi|160047	p101/acidic basic repeat antigen (Plasmodium falciparum) pir|A29232|A29232 101K malaria antigen precursor - Plasmodium falciparum (strain Camp)	61	35	822
168	3	1915	2361	gi|1499694	HIT protein, member of the HIT-family (Methanococcus jannaschii)	61	41	447
171	9	9675	7948	gi|467436	similar to SpoVB (Bacillus subtilis)	61	38	1728
174	3	1042	2340	gi|216374	glutaryl 7-ACA acylase precursor (Bacillus laterosporus)	61	49	1299
190	4	5034	4111	gi|409286	bmrU (Bacillus subtilis)	61	37	924
216	1	2	190	gi|415861	eukaryotic initiation factor 2 beta (eIF-2 beta) (Oryctolagus uniculus)	61	29	189
227	7	4161	5048	gi|216341	ORF for methionine amino peptidase (Bacillus subtilis)	61	41	888
238	4	1959	3047	gi|409543	CbrC protein (Erwinia chrysanthemi)	61	38	1089
247	1	2	694	gi|537231	ORF_f579 (Escherichia coli)	61	38	693
247	2	678	1014	gi|142226	chvD protein (Agrobacterium tumefaciens)	61	40	357
257	2	3523	2627	gi|699379	glvr-1 protein (Mycobacterium leprae)	61	40	897
268	2	3419	3051	gi|40364	ORFA1 (Clostridium acetobutylicum)	61	41	369
275	4	4621	4827	gi|1204848	hypothetical protein (GP:H87049_57) (Haemophilus influenzae)	61	36	207
277	1	1	1845	gi|7H4897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - Streptococcus pneumoniae	61	45	1845
278	9	8003	7032	gi|467462	cysteine synthetase A (Bacillus subtilis)	61	43	972
278	10	9878	8535	gi|1205919	Na+ and Cl- dependent gamma-aminobutyric acid transporter (Haemophilus influenzae)	61	38	1344
283	1	1	366	gi|755607	polyA polymerase (Bacillus subtilis)	61	36	366
288	2	1918	1496	gi|388108	cell wall enzyme (Enterococcus faecalis)	61	43	423
291	1	86	334	gi|454265	FBP3 (Petunia hybrida)	61	38	249
318	1	1104	694	gi|290531	similar to beta-glucoside transport protein (Escherichia coli) sp|P11651|PTB_ECOLI PTS SYSTEM, ARBUTIN-LIKE IIB COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT) (EC 2.7.1.69).	61	47	411
330	2	1912	1190	gi|1001805	hypothetical protein (Synechocystis sp.)	61	41	723

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
385	2	1513	1025	gi|533098	DnaD protein [Bacillus subtilis]	61	42	489
426	1	794	399	gi|1303853	YqgF [Bacillus subtilis]	61	44	396
438	3	810	1421	gi|1293660	AbaA2 [Streptomyces coelicolor]	61	36	612
454	1	1580	792	gi|733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostelium iscoideum]	61	30	789
464	2	784	560	gi|1121120	C53B7.5 gene product [Caenorhabditis elegans]	61	38*	225
470	8	6077	7357	gi|623073	ORF360; putative [Bacteriophage LL-H]	61	47	1281
509	1	554	279	gi|467484	unknown [Bacillus subtilis]	61	45	276
555	3	1916	1296	gi|141800	anthranilate synthase glutamine amidotransferase [Acinetobacter alcoaceticus]	61	42	621
589	1	1711	857	gi|467090	B2235_C2_195 [Mycobacterium leprae]	61	47	855
585	2	961	803	sp|P16686|SURE_	SURVIVAL PROTEIN SURE HOMOLOG (FRAGMENT)	61	33	159
592	3	1694	1422	gi|1221602	immunity repressor protein [Haemophilus influenzae]	61	32	273
603	1	43	357	gi|507738	llap [Vibrio parahaemolyticus]	61	33	315
669	1	2467	1235	gi|1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ....; putative [Bacillus subtilis]	61	37	1233
675	3	805	1101	gi|403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir|S17251|S17251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	61	36	297
703	1	1656	829	gi|537181	ORF_f470 [Escherichia coli]	61	32	828
728	1	1628	816	gi|806281	DNA polymerase I [Bacillus stearothermophilus]	61	39	813
821	1	61	318	gi|709992	hypothetical protein [Bacillus subtilis]	61	38	258
856	2	2313	1567	gi|609310	portal protein gp3 [Bacteriophage HK97]	61	40	747
921	1	1081	542	gi|143213	putative [Bacillus subtilis]	61	38	540
1124	1	59	370	gi|1107541	C31D9.8 [Caenorhabditis elegans]	61	26	312
1492	1	548	276	gi|406397	unknown [Mycoplasma genitalium]	61	32	273
1602	1	46	318	gi|733522	phosphatidylinositol-4,5-diphosphate 3-kinase [Dictyostelium iscoideum]	61	34	273
2500	1	577	290	gi|1045964	hypothetical protein (GB:U14001_297) [Mycoplasma genitalium]	61	31	288
2968	1	2	808	gi|397526	clumping factor [Staphylococcus aureus]	61	55	807
3076	1	3	248	gi|149373	ORF 1 [Lactococcus lactis]	61	41	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1609	2	207	401	gi|1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus (Bacillus subtilis)	61	39	195
1662	1	1477	740	gi|1303813	YgdH (Bacillus subtilis)	61	42	738
1672	1	2	442	gi|84897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	441
1724	1	2	220	gi|1009366	Respiratory nitrate reductase (Bacillus subtilis)	61	41	219
1728	1	3	398	gi|677943	AppD (Bacillus subtilis)	61	46	396
1884	1	3	401	gi|784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	47	399
1971	1	3	383	gi|784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	45	381
4038	1	661	359	gi|1133399:n	large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)	61	24	303
4041	1	546	274	gi|1413953	ipa-29d gene product (Bacillus subtilis)	61	48	273
4047	1	1	402	gi|528991	unknown (Bacillus subtilis)	61	42	402
4102	1	1	345	gi|976025	HtrA (Escherichia coli)	61	46	345
4155	1	1	336	gi|784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	336
4268	1	463	233	gi|450688	hadM gene of Ecoprrl gene product (Escherichia coli) pir|S38437|S38437 hadM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	61	38	231
4374	1	542	273	gi|784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	50	270
4389	1	2	172	gi|147516	ribokinase (Escherichia coli)	61	35	171
4621	1	2	268	gi|784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir|A56390|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae	61	47	267
4663	1	27	227	gi|976025	HtrA (Escherichia coli)	61	50	201
4	6	6663	5536	gi|1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus (Bacillus subtilis)	60	43	1128

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
11	6	3426	3725	gi|410748	ring-infested erythrocyte surface antigen (Plasmodium falciparum) pir|A25526|A25526 ring-infected erythrocyte surface antigen recursor - Plasmodium falciparum (strain FC27/Papua New Guinea) sp|P13830|RESA_PLAFF RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN R2	60	24	300
11	14	11035	10313	gi|1217851	carbonyl reductase (NADPH) [Rattus norvegicus]	60	28	723
16	12	11917	12930	gi|1001451	hypothetical protein [Synechocystis sp.]	60	37	1014
33	1	26	469	gi|388109	regulatory protein [Enterococcus faecalis]	60	41	444
37	13	10914	9834	gi|1336656	Orf1 [Bacillus subtilis]	60	40	981
39	4	4364	4522	gi|4872	ORF 4 [Saccharomyces kluyveri]	60	47	159
41	1	2047	1025	gi|142822	D-alanine racemase cds [Bacillus subtilis]	60	39	1023
43	4	2474	3607	gi|468046	para-nitrobenzyl esterase [Bacillus subtilis]	60	40	1134
44	10	6756	7769	gi|441234	thiF [Escherichia coli]	60	52	1014
45	10	8874	9074	gi|343949	(var1(40.0) [Saccharomyces cerevisiae]	60	44	201
56	18	27842	26430	gi|468764	mocR gene product [Rhizobium meliloti]	60	35	1413
60	2	173	388	gi|1303864	YqgQ [Bacillus subtilis]	60	33	216
63	2	357	1619	gi|467124	ureD: B229_C3_234 [Mycobacterium leprae]	60	43	1263
69	1	787	395	gi|1518853	OdfA [Salmonella typhimurium]	60	16	193
88	1	1	1188	gi|140429	putative transcriptional regulator [Bacillus stearothermophilus]	60	30	1188
92	6	4735	3881	gi|349227	transmembrane protein [Escherichia coli]	60	37	855
92	7	5996	4923	gi|466613	nlkB [Escherichia coli]	60	38	1074
93	1	949	476	gi|1510925	coenzyme F420-reducing hydrogenase, beta subunit [Methanococcus jannaschii]	60	27	474
96	6	7366	7578	gi|472715	accessory protein [Arthrobacter picalrola]	60	30	213
98	6	3212	4069	gi|467425	unknown [Bacillus subtilis]	60	42	858
102	10	7158	7430	gi|143092	acetolactate synthase small subunit [Bacillus subtilis] sp|P37253|ILVN_BACSU ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC .1.3.18) (AHAS) (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT) (ALS)	60	37	273
109	11	9127	10515	gi|1255259	o-succinylbenzoic acid (OSB) CoA ligase [Staphylococcus aureus]	60	28	1389
109	12	10499	11656	gi|141954	beta-ketothiolase [Alcaligenes eutrophus]	60	41	1158
119	2	4630	3134	gi|1524280	unknown [Mycobacterium tuberculosis]	60	45	1497

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
121	9	6957	7646	|gi|1107529	ceuC gene product [Campylobacter coli]	60	35	690
140	7	7704	6013	|gi|146547	kdpA [Escherichia coli]	60	45	1692
145	1	2	703	|gi|1460077	unknown [Mycobacterium tuberculosis]	60	23	702
150	3	2809	2216	|gi|1146230	putative [Bacillus subtilis]	60	40	594
157	2	1389	961	|gi|1303975	YqjX [Bacillus subtilis]	60	30'	429
158	5	5125	4769	|gi|1449288	unknown [Mycobacterium tuberculosis]	60	36	357
159	1	511	257	|gi|580912	murD gene product [Bacillus subtilis]	60	43	255
160	1	159	1187	|gi|1204532	hypothetical protein (GB:L19201_29) [Haemophilus influenzae]	60	14	1029
161	14	8249	7866	|gi|1496003	[ORFJ; PepV; putative oligoendopeptidase based on homology with Lactococcus lactis PepF (GenBank Accession Number Z32522) [Caldicellulosiruptor saccharolyticus]	60	34	384
172	3	1331	2110	|gi|485280	28.2 kDa protein [Streptococcus pneumoniae]	60	33	780
173	2	4082	2460	|gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	60	41	1623
173	4	5963	4953	|gi|1100737	NADP dependent leukotriene b4 12-hydroxydehydrogenase [Sus scrofa]	60	44	1011
198	1	3	995	|gi|413943	lpa-19d gene product [Bacillus subtilis]	60	42	993
201	4	3641	4573	|sp|P37028|YADT_	HYPOTHETICAL 29.4 KD PROTEIN IN HEXL-PFS INTERGENIC REGION PRECURSOR.	60	37	933
203	3	3269	2415	|gi|927798	D9719.34p; CAI: 0.14 [Saccharomyces cerevisiae]	60	43	895
206	9	12234	12515	|sp|P37347|YECO_	HYPOTHETICAL 21.8 KD PROTEIN IN ASPS 5'REGION.	60	47	282
212	4	1213	1410	|gi|332711	hemagglutinin-neuraminidase fusion protein [Human parainfluenza irus 3]	60	34	198
214	1	65	1153	|gi|1204366	hypothetical protein (GB:U14003_130) [Haemophilus influenzae]	60	36	1089
237	1	2	937	|gi|149377	HlsD [Lactococcus lactis]	60	40	936
241	6	5696	4998	|gi|1046160	hypothetical protein (GB:U00021_5) [Mycoplasma genitalium]	60	37	699
260	6	5919	6485	|gi|431950	similar to a B.subtilis gene (GB: BACHEMRHY_5) [Clostridium asteurianum]	60	35	567
264	1	2432	1218	|gi|397526	clumping factor [Staphylococcus aureus]	60	53	1215
267	1	3	1409	|gi|148316	NaH-antiporter protein [Enterococcus hirae]	60	27	1407
275	3	3804	4595	|pir|F36A89|F368	leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis [strain IL1403]	60	35	792
291	3	860	1198	|gi|1208689	coded for by C. elegans cDNA yk130e12.5; contains C2H2-type zinc fingers [Caenorhabditis elegans]	60	33	339

TABLE 2

134

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	# aa	% ident	length (nt)
307	6	3421	3176	gi|1070014	protein-dependent (Bacillus subtilis)	60	36	246
316	8	4957	5823	gi|413952	ipa-28d gene product (Bacillus subtilis)	60	41	867
328	4	2996	3484	gi|1204484	membrane-associated component, branched amino acid transport system [Haemophilus influenzae]	60	39	489
332	5	4887	4363	gi|1205449	colicin V production protein (pur regulon) [Haemophilus influenzae]	60	37	525
357	1	1062	532	gi|A87842	single-stranded DNA-specific exonuclease [Escherichia coli]	60	41	531
375	2	96	362	gi|1057	adenylyl cyclase gene product (Saccharomyces kluyveri) r[JQ1145|OYBYK adenylate cyclase (EC 4.6.1.1) - yeast ccharomyces kluyveri)	60	47	267
397	1	66	416	gi|709999	Glucarate dehydratase [Bacillus subtilis]	60	37	351
409	1	2	163	gi|499700	glycogen phosphorylase [Saccharomyces cerevisiae]	60	35	162
453	4	914	1237	gi|1196899	unknown protein [Staphylococcus aureus]	60	36	324
453	7	3838	3620	sp|P12222|YCF1_	HYPOTHETICAL 226 KD PROTEIN (ORF 1901)	60	31	219
470	2	622	945	pir|S10782|S307	integrin homolog - yeast (Saccharomyces cerevisiae)	60	31	324
500	1	118	606	gi|467407	unknown [Bacillus subtilis]	60	36	489
503	3	752	982	gi|167835	myosin heavy chain [Dictyostelium discoideum]	60	34	231
505	4	2238	3563	gi|1510732	NADH oxidase [Methanococcus jannaschii]	60	26	1326
523	1	3	1043	gi|143331	alkaline phosphatase regulatory protein (Bacillus subtilis) pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P23545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN IIOR (EC 2.7.3.-).	60	41	1041
543	1	1	465	gi|1511103	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	60	40	465
545	1	1	726	gi|1498192	putative [Pseudomonas aeruginosa]	60	40	726
556	1	2	1054	gi|1477402	tex gene product [Bordetella pertussis]	60	42	1053
578	1	974	489	gi|1205129	H. influenzae predicted coding region HI0882 [Haemophilus influenzae]	60	42	486
594	1	1	624	gi|1212755	adenylyl cyclase [Aeromonas hydrophila]	60	45	624
604	1	3	530	gi|145925	fecB [Escherichia coli]	60	42	528
620	1	926	465	gi|1205483	bicyclomycin resistance protein [Haemophilus influenzae]	60	33	462
630	2	871	1122	gi|1486242	unknown [Bacillus subtilis]	60	41	252
645	2	574	425	gi|1205136	serine hydroxymethyltransferase (serine methylase) [Haemophilus influenzae]	60	28	150

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
684	1	1082	843	gi|1205538	hypothetical protein (GB:U14003_302) [Haemophilus influenzae]	60	39	240
786	1	967	485	gi|1402944	orfRM1 gene product [Bacillus subtilis]	60	46	483
844	1	588	346	gi|790943	urea amidolyase [Bacillus subtilis]	60	40	243
851	1	1	726	gi|159661	GMP reductase [Ascaris lumbricoides]	60	41	726
871	1	1746	874	gi|1001493	hypothetical protein [Synechocystis sp.]	60	39	873
896	1	1558	839	gi|604926	NADH dehydrogenase, subunit 5 [Schizophylum commune] sp|P50168|NU5M_SCHCO NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC .6.5.3).	60	39	720
908	2	448	753	gi|662880	novel hemolytic factor [Bacillus cereus]	60	31	306
979	1	2	595	gi|1429255	putative; orf1 [Bacillus subtilis]	60	30	594
1078	1	669	502	gi|561055	inner membrane copper tolerance protein [Escherichia coli] gi|871029 disulphide isomerase like protein [Escherichia coli] pir|647295|647295 inner membrane copper tolerance protein - escherichia coli	60	40	168
1112	1	1150	620	gi|407885	ORF3 [Streptomyces griseus]	60	34	531
1135	1	484	275	gi|1171407	vpsRp [Saccharomyces cerevisiae]	60	36	210
1146	1	17	562	gi|1239981	hypothetical protein [Bacillus subtilis]	60	36	546
1291	1	716	360	pir|S57530|S575	carboxyl esterase - Acinetobacter calcoaceticus	60	30	357
1312	1	336	169	gi|1222056	aminotransferase [Haemophilus influenzae]	60	44	168
1429	1	3	146	gi|1205619	ferritin like protein [Haemophilus influenzae]	60	39	144
1722	1	570	286	gi|240052	dihydroflavonol-4-reductase, DFR [Hordeum vulgare=barley, cv. Gula, eptide, 354 aa]	60	36	285
2350	1	385	200	gi|497626	ORF 1 [Plasmid pAQ1]	60	20	186
2936	1	519	310	gi|508981	prephenate dehydratase [Bacillus subtilis]	60	48	210
3027	1	568	302	gi|1146199	putative [Bacillus subtilis]	60	37	267
3044	1	20	208	gi|1407784	orf-1; novel antigen [Staphylococcus aureus]	60	51	189
3155	1	2	226	gi|1046097	cytadherence-accessory protein [Mycoplasma genitalium]	60	34	225
3603	1	368	186	gi|510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA ehydrogenase alpha-subunit [Rattus norvegicus]	60	42	183
3665	1	486	244	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	60	42	243
3747	1	3	146	gi|474192	iucC gene product [Escherichia coli]	60	36	144

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3912	1	3	335	gi|488695	novel antigen: orf-2 [Staphylococcus aureus]	60	44	333
4072	1	3	272	gi|405879	yelH [Escherichia coli]	60	33	270
4134	1	510	352	gi|780656	chemoreceptor protein [Rhizobium leguminosarum bv. viciae] gi|780656 chemoreceptor protein [Rhizobium leguminosarum bv. iclae]	60	28	159
4207	2	677	402	gi|602031	similar to trimethylamine DH [Mycoplasma capricolum] pir|S49950|S49950 probable trimethylamine dehydrogenase (EC .5.99.7) - Mycoplasma capricolum (SGC3) (fragment)	60	41	276
4243	1	127	324	gi|899317	peptide synthetase module [Microcystis aeruginosa] pir|S49111|S49311 probable amino acid activating domain - Microcystis aeruginosa (fragment) (SUB 144-528)	60	42	198
4110	1	624	313	gi|308980	phaB (Bacillus subtilis]	60	28	312
4145	1	343	173	gi|510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA ehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
4382	1	498	280	gi|47382	acyl-CoA-dehydrogenase [Streptomyces purpurascens]	60	48	219
4474	1	53	223	gi|510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA ehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
23	4	4518	3523	gi|426446	VipB protein [Salmonella typhi]	59	39	996
33	2	707	1483	pir|S48604|S486	hypothetical protein - Mycoplasma capricolum (SGC3) (fragment)	59	33	777
33	5	4651	5853	gi|6721	F59B2.3 [Caenorhabditis elegans]	59	33	1203
17	2	1228	2298	gi|142833	ORF2 [Bacillus subtilis]	59	37	910
38	21	16784	16593	gi|912576	BiP [Phaeodactylum tricornutum]	59	40	192
52	3	2648	2349	gi|536972	ORF_o90a [Escherichia coli]	59	44	300
54	12	14181	13402	gi|489940	transcription regulator [Bacillus subtilis]	59	37	780
57	3	4397	3339	gi|508176	[Gat-1-P-DH, NAD dependent [Escherichia coli]	59	40	1059
66	1	986	495	gi|1303901	Yqhf [Bacillus subtilis]	59	34	492
67	7	6552	7460	gi|912461	nlkC [Escherichia coli]	59	37	909
70	7	5383	6365	gi|1399822	PhoD precursor (Rhizobium meliloti]	59	46	984
78	1	1	1449	gi|971345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp|P45860|YWIE_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA INTERGENIC REGION.	59	39	1449
82	10	14329	15534	gi|490328	LORF F (unidentified)	59	44	1206

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
89	2	1602	958	gi|642801	unknown (Saccharomyces cerevisiae)	59	32	645
96	4	4940	5473	gi|1333802	protein of unknown function (Rhodobacter capsulatus)	59	33	534
98	1	2	820	gi|467421	similar to B. subtilis DnaH (Bacillus subtilis)	59	34	819
119	1	166	1557	gi|143122	ORF B; putative (Bacillus firmus)	59	36	1392
120	10	6214	6756	gi|15354	ORF 55.9 (Bacteriophage T4)	59	39	543
120	16	12476	13510	gi|1086575	BotA (Rhizobium meliloti)	59	44	1035
123	1	386	195	gi|984737	catalase (Campylobacter jejuni)	59	38	192
130	1	370	645	gi|1256634	25.8% identity over 120 aa with the Synenococcus sp. MpeV protein; putative (Bacillus subtilis)	59	31	276
131	4	5278	5712	gi|1510655	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	59	39	435
164	1	3	509	gi|1001342	hypothetical protein (Synechocystis sp.)	59	41	507
164	4	1529	2821	gi|1205165	hypothetical protein (SP:P37764) (Haemophilus influenzae)	59	35	1293
164	19	19643	21376	gi|1001381	hypothetical protein (Synechocystis sp.)	59	34	1734
173	3	4727	3717	gi|1184121	auxin-induced protein (Vigna radiata)	59	50	1011
179	2	2218	1688	gi|143036	unidentified gene product (Bacillus subtilis)	59	33	531
195	12	12669	11503	gi|762778	NifS gene product (Anabaena azollae)	59	41	1167
201	5	4702	5670	gi|1510240	hemin permease (Methanococcus jannaschii)	59	32	969
201	7	5719	6315	gi|1511456	M. jannaschii predicted coding region MJ1437 (Methanococcus jannaschii)	59	34	597
209	1	102	461	gi|1204666	hypothetical protein (GB:X73124_53) (Haemophilus influenzae)	59	42	360
214	3	1050	2234	gi|551531	2-nitropropane dioxygenase (Williopsis saturnus)	59	36	1185
214	5	3293	4135	gi|1303709	YrkJ (Bacillus subtilis)	59	32	843
217	2	3381	2167	gi|1290489	dtp (CG Site No. 18430) (Escherichia coli)	59	44	1215
237	5	3078	3785	gi|149382	HisA (Lactococcus lactis)	59	38	708
251	2	376	960	gi|1303791	YqoJ (Bacillus subtilis)	59	34	585
286	1	1621	812	gi|46551	transmembrane protein (kdpD) (Escherichia coli)	59	31	810
316	5	4978	3860	gi|405879	yeiH (Escherichia coli)	59	32	1119
370	3	600	761	gi|1303794	YqeM (Bacillus subtilis)	59	35	162

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	start (nt)	stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
382	1	1009	506	g|547513	orf3 [Haemophilus influenzae]	59	34	504
391	3	1620	1273	g|152901	ORF 3 (Spirochaeta aurantia)	59	37	348
406	3	2805	1705	g|709992	hypothetical protein (Bacillus subtilis)	59	34	1101
426	5	3802	3245	g|1204610	iron(III) dicitrate transport ATP-binding protein FECE [Haemophilus influenzae]	59	36	558
429	2	1513	1148	g|1064809	homologous to sp:HTRA_ECOLI (Bacillus subtilis)	59	42	366
460	2	708	1301	g|466882	ppsi; B1496_C2_189 (Mycobacterium leprae)	59	37	594
461	4	2212	3135	g|1498295	homoserine kinase homolog (Streptococcus pneumoniae)	59	37	924
473	1	2929	1607	g|147989	trigger factor (Escherichia coli)	59	40	1323
480	8	5862	6110	g|1205311	(3R)-hydroxymyristol acyl carrier protein dehydrase [Haemophilus influenzae]	59	40	249
521	1	14	1354	pir|A25620|A256	staphylocoagulase - Staphylococcus aureus (fragment)	59	32	1341
534	4	2994	4073	g|153746	mannitol-phosphate dehydrogenase (Streptococcus mutans) pir|C44798|C44798 mannitol-phosphate dehydrogenase MtlD - streptococcus mutans	59	36	1080
535	1	1	954	g|1469939	group B oligopeptidase PepB (Streptococcus agalactiae)	59	33	954
551	3	2836	3186	g|1204511	bacterioferritin comigratory protein [Haemophilus influenzae]	59	45	351
573	2	449	940	g|386681	ORF YAL022 (Saccharomyces cerevisiae)	59	36	492
650	1	5	748	g|396400	similar to eukaryotic Na/H+ exchangers [Escherichia coli] sp|P32703|YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC REGION (O549).	59	30	744
664	1	566	285	g|1262748	LukF-PV like component (Staphylococcus aureus)	59	33	282
670	1	3	455	g|1122758	unknown (Bacillus subtilis)	59	42	453
674	3	543	929	g|293033	integrase [Bacteriophage phi-LC3]	59	46	387
758	1	349	176	g|1500472	M. jannaschii predicted coding region MJ1577 (Methanococcus jannaschii)	59	37	174
771	2	2270	1461	g|522150	bromoperoxidase BPO-A1 [Streptomyces aureofaciens] sp|P39912|BPA1_STRAU NON-HAEM BROMOPEROXIDASE BPO-A1 (EC 1.11.1.-) BROMIDE PEROXIDASE) (BPO1). (SUB 2-275)	59	44	810
825	1	2191	1097	g|397526	clumping factor (Staphylococcus aureus)	59	47	1095
1052	2	1094	723	g|289262	comE ORF3 (Bacillus subtilis)	59	36	372
1152	1	373	188	g|1276668	ORF238 gene product (Porphyra purpurea)	59	37	186

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
119A	1	492	247	gi|142439	ATP-dependent nuclease [Bacillus subtilis]	59	26	246
1441	1	468	235	gi|1045942	glycyl-tRNA synthetase [Mycoplasma genitalium]	59	37	234
2103	1	1	186	gi|459250	triacylglycerol lipase [Galactomyces geotrichum]	59	33	186
2205	1	793	398	gi|1303794	YqeM [Bacillus subtilis]	59	38	396
2578	1	484	284	gi|258003	insulin-like growth factor binding protein complex acid-labile ubunit (rats, liver, Peptide, 603 aa)	59	48	201
2967	2	145	348	gi|1212730	YqhK [Bacillus subtilis]	59	44	204
3012	1	3	248	gi|773571	neurofilament protein NF70 [Helix aspersa]	59	31	246
3544	1	3	401	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3548	1	3	401	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3580	1	698	351	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	348
3720	1	722	363	gi|1408494	homologous to penicillin acylase [Bacillus subtilis]	59	36	360
4171	1	3	226	gi|1055218	crotonase [Clostridium acetobutylicum]	59	42	294
4305	1	618	310	gi|1524193	unknown [Mycobacterium tuberculosis]	59	39	109
1R	1	1242	622	gi|146913	N-acetylglucosamine transport protein [Escherichia coli pir[a29895]WQBC2N phosphotransferase system enzyme II (EC 7.1.69), N-acetylglucosamine-specific - Escherichia coli sp|P09323|PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC OMPONENT (EIIA	58	43	621
20	7	7020	5845	gi|50502	collagen alpha chain precursor (AA -27 to 1127) [Mus musculus]	58	50	1176
21	5	3234	3626	gi|1054860	phosphoribosyl anthranilate isomerase [Thermotoga maritima]	58	32	393
23	2	2861	1669	gi|1276880	EpsG [Streptococcus thermophilus]	58	29	1173
23	10	9301	8090	pir|A31133|A311	diaminopimelate decarboxylase (EC 4.1.1.20) - Pseudomonas aeruginosa	58	37	1212
18	29	22555	22884	gi|973249	vestitone reductase [Medicago sativa]	58	37	330
44	1	2	406	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	58	33	405
45	1	1	552	gi|29464	embryonic myosin heavy chain (1085 AA) [Homo sapiens] tr|S12460|S12460 myosin beta heavy chain - human	58	33	552
55	2	759	538	gi|158852	glucose regulated protein [Echinococcus multilocularis]	58	32	222
62	13	8493	8068	gi|975353	xinase-associated protein B [Bacillus subtilis]	58	35	426
61	3	1553	1717	gi|166926	(Arabidopsis thaliana unidentified mRNA sequence, complete cds.), ane product [Arabidopsis thaliana]	58	35	165

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	start (nt)	stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
67	13	12017	11229	gi|1228083	[NADH dehydrogenase subunit 2 (Chorthippus parallelus)]	58	41	789
96	8	8208	9167	gi|709992	[hypothetical protein (Bacillus subtilis)]	58	42	960
107	2	2065	1364	gi|806327	[Escherichia coli hrpA gene for A protein similar to yeast PRP16 and RP22 (Escherichia coli)]	58	37	702
112	7	4519	5613	gi|155588	[glucose-fructose oxidoreductase (Zymomonas mobilis) pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis]	58	38,	1095
114	6	731R	6503	gi|1377843	[unknown (Bacillus subtilis)]	58	38	816
143	2	2261	1395	pir|A56605|A56	[mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum]	58	31	867
151	2	717	950	gi|1370261	[unknown (Mycobacterium tuberculosis)]	58	31	234
154	6	6015	4627	gi|1209277	[pCTHom1 gene product (Chlamydia trachomatis)]	58	41	1389
154	16	14281	13541	gi|146613	[DNA ligase (EC 6.5.1.2) (Escherichia coli)]	58	39	741
155	3	2269	1892	gi|1303917	[YqiB (Bacillus subtilis)]	58	34	378
174	1	1056	529	gi|90419R	[hypothetical protein (Bacillus subtilis)]	58	26	528
189	4	1533	1769	gi|467383	[DNA binding protein (probable) (Bacillus subtilis)]	58	25	237
201	3	2669	3307	gi|1511453	[endonuclease III (Methanococcus jannaschii)]	58	34	639
20R	1	2	238	gi|1276729	[phycobilisome linker polypeptide (Porphyra purpurea)]	58	29	237
220	11	14575	13058	gi|397526	[clumping factor (Staphylococcus aureus)]	58	51	1518
231	3	1629	1474	gi|1002520	[HueS (Bacillus subtilis)]	58	45	156
233	6	4201	3497	gi|1463023	[No definition line found (Caenorhabditis elegans)]	58	39	705
243	10	9303	10082	gi|537207	[ORF_f277 (Escherichia coli)]	58	32	780
257	1	331	1143	gi|1340128	[ORF1 (Staphylococcus aureus)]	58	44	813
302	2	460	801	gi|40174	[ORF X (Bacillus subtilis)]	58	34	342
307	11	6984	6127	gi|1303842	[YqfU (Bacillus subtilis)]	58	30	858
321	3	1914	2747	gi|1239996	[hypothetical protein (Bacillus subtilis)]	58	41	834
342	4	2724	3497	gi|1454838	[ORF 6; putative (Pseudomonas aeruginosa)]	58	41	774
348	1	1	663	gi|467478	[unknown (Bacillus subtilis)]	58	36	663
401	2	384	605	gi|143407	[para-aminobenzoic acid synthase, component I (pab) (Bacillus subtilis)]	58	53	222

TABLE 2

141

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
437	1	325	1554	gi|1301866	YrgS [Bacillus subtilis]	58	35	1230
445	1	105	1442	gi|581583	protein A (Staphylococcus aureus)	58	22	1338
453	3	789	965	gi|1009455	unknown (Schizosaccharomyces pombe)	58	34	177
453	5	2746	2047	gi|537214	yjjG gene product (Escherichia coli)	58	40	702
479	2	731	1444	gi|1256621	26.7% of identity.in 165 aa to a Thermophilic bacterium hypothetical protein 6: putative (Bacillus subtilis)	58	36	714
490	1	909	547	gi|580920	rodD (gtcA) polypeptide (AA 1-67) (Bacillus subtilis) pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13484|TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROTEIN B).	58	16	363
517	1	1	1164	br|P47264|Y018_	HYPOTHETICAL HELICASE MG018.	58	30	1164
517	6	4182	4544	gi|453422	orf269 gene product (Mycoplasma hominis)	58	29	363
546	3	2802	4019	gi|886052	restriction modification system S subunit (Spiroplasma citri) gi|886052 restriction modification system S subunit (Spiroplasma ltri)	58	37	1218
562	1	3	179	gi|43831	nifS protein (AA 1-409) (Klebsiella pneumoniae)	58	34	177
600	2	1347	1156	gi|1181839	unknown (Pseudomonas aeruginosa)	58	48	192
604	2	1231	1001	gi|1001353	hypothetical protein (Synechocystis sp.)	58	41	231
619	1	1	504	gi|903748	integral membrane protein [Homo sapiens]	58	43	504
625	1	2	364	gi|1208474	hypothetical protein (Synechocystis sp.)	58	41	363
635	1	1492	755	gi|1510995	transaldolase [Methanococcus jannaschii]	58	41	738
645	1	1	846	gi|677882	[ileal] sodium-dependent bile acid transporter [Rattus norvegicus] gi|677882 ileal sodium-dependent bile acid transporter [Rattus orvegicus]	58	33	846
645	3	906	1556	gi|1239999	hypothetical protein (Bacillus subtilis)	58	41	651
665	1	771	532	gi|1204262	hypothetical protein (GB:L10328_61) (Haemophilus influenzae)	58	39	240
674	1	635	327	gi|498817	ORF8; homologous to small subunit of phage terminases (Bacillus ubtilis)	58	39	309
675	2	1312	806	gi|42181	osmC gene product (Escherichia coli)	58	28	507
745	1	618	310	gi|1205422	coenzyme PQQ synthesis protein III (pqqIII) (Haemophilus influenzae)	58	32	309
799	2	242	1174	gi|1204669	collagenase (Haemophilus influenzae)	58	16	933
800	2	1096	614	gi|171963	tRNA isopentenyl transferase (Saccharomyces cerevisiae) sp|P07884|MOD5_YEAST TRNA ISOPENTENYLTRANSFERASE (EC 2.5.1.8) ISOPENTENYL- DIPHOSPHATE: TRNA ISOPENTENYLTRANSFERASE) (IPP RANSFERASE) (IPPT).	58	37	483

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
854	1	1108	605	gi|466778	lysine specific permease [Escherichia coli]	58	44	504
885	1	481	242	gi|861199	protoporphyrin IX Mg-chelatase subunit precursor [Hordeus vulgare]	58	33	240
891	1	3	527	gi|1293660	[AbsA2 (Streptomyces coelicolor)	58	31	525
942	1	931	467	gi|405567	traH [Plasmid pSK41]	58	30	465
1002	1	952	521	gi|577649	proLUKH [Staphylococcus aureus]	58	34*	432
1438	1	1	261	gi|581558	isoleucyl tRNA synthetase [Staphylococcus aureus] sp|P41368|SYIP_STAAU ISOLEUCYL-TRNA SYNTHETASE, HUPIROCIN RESISTANT EC 6.1.1.5) (ISOLEUCINE--TRNA LIGASE) (ILERS) (HUPIROCIN RESISTANCE ROTEIN).	58	30	261
1442	1	2	463	gi|971394	similar to Acc.No. D26185 [Escherichia coli]	58	34	462
1873	1	480	241	gi|1339951	small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	58	38	240
1876	1	3	158	gi|529216	No definition line found [Caenorhabditis elegans] sp|P46503|YLX7_CAEEL HYPOTHETICAL 7.3 KD PROTEIN F23F12.7 IN HROMOSOME III.	58	33	156
1989	1	108	401	gi|1405458	YneR [Bacillus subtilis]	58	29	294
2109	1	3	401	gi|1001801	hypothetical protein [Synechocystis sp.]	58	31	399
2473	1	288	145	gi|510140	oligoendopeptidase F [Lactococcus lactis]	58	38	144
2523	1	452	228	gi|644873	catabolic dehydroquinate dehydratase [Acinetobacter calcoaceticus]	58	37	225
3041	1	2	211	gi|1205367	oligopeptide transport ATP-binding protein [Haemophilus influenzae]	58	39	210
3094	1	3	263	gi|1118528	isochorismate synthase [Bacillus subtilis]	58	38	261
3706	1	3	383	gi|456614	mevalonate kinase [Arabidopsis thaliana]	58	44	381
3854	1	1	402	gi|808869	human gcp372 [Homo sapiens]	58	32	402
4042	1	51	224	gi|508551	ribulosa-1,5 bisphosphate carboxylase large subunit -methyltransferase [Pisum sativum]	58	37	174
4278	1	3	206	gi|180189	cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi|182737 cerebellar degeneration-associated protein [Homo sapiens] pir|A29770|A29770 cerebellar degeneration-related protein - human	58	37	204
19	7	7818	7363	gi|1001516	hypothetical protein [Synechocystis sp.]	57	31	456
23	11	9663	8872	gi|606066	ORF_f256 [Escherichia coli]	57	29	792
31	1	4801	2402	gi|153146	ORF3 [Streptomyces coelicolor]	57	32	2400
38	14	11611	10796	gi|144859	ORF B [Clostridium perfringens]	57	31	816
46	14	12063	13046	gi|1001319	hypothetical protein [Synechocystis sp.]	57	25	984

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
51	3	1411	1187	pir|B33856|B338	hypothetical 80K protein - Bacillus sphaericus	57	38	225
54	1	1	453	gi|6R4950	staphylococcal accessory regulator A [Staphylococcus aureus]	57	31	453
75	1	3	239	gi|1000470	C27B7.7 [Caenorhabditis elegans]	57	42	237
92	5	3855	3061	gi|143607	sporulation protein [Bacillus subtilis]	57	35	795
96	3	4006	4773	gi|144297	acetyl esterase (XynC) [Caldocellum saccharolyticum] pir|B37202|B37202 acetylesterase (EC 3.1.1.6) (XynC) - Caldocellum saccharolyticum	57	34	768
107	3	1480	2076	gi|460955	TagS [Vibrio cholerae]	57	42	597
109	8	5340	5933	gi|1438846	unknown [Bacillus subtilis]	57	41	594
112	9	6679	7701	gi|486250	unknown [Bacillus subtilis]	57	33	1023
114	4	6384	4108	gi|871456	putative alpha subunit of formate dehydrogenase [Methanobacterium thermoautotrophicum]	57	37	2277
126	2	430	1053	gi|288301	ORF2 gene product [Bacillus megaterium]	57	37	624
131	5	6537	6277	gi|1511160	M. jannaschii predicted coding region MJ1163 [Methanococcus jannaschii]	57	38	261
133	3	2668	2201	gi|1303912	YqhW [Bacillus subtilis]	57	40	468
133	4	3383	2784	gi|1221884	(urea?) amidolyase [Haemophilus influenzae]	57	37	600
147	4	2164	1694	gi|467469	unknown [Bacillus subtilis]	57	33	471
160	2	1293	1060	gi|558604	chitin synthase 2 [Neurospora crassa]	57	28	234
163	8	5687	4764	gi|145580	rarD gene product [Escherichia coli]	57	38	924
168	6	4336	5325	gi|39782	33kDa lipoprotein [Bacillus subtilis]	57	32	990
170	5	3297	3455	gi|603404	Yer164p [Saccharomyces cerevisiae]	57	37	159
221	6	8026	6809	gi|1136221	carboxypeptidase [Sulfolobus solfataricus]	57	32	1218
226	3	1348	1791	gi|1288969	fibronectin binding protein [Streptococcus dysgalactiae] pir|S33850|S33850 fibronectin-binding protein - Streptococcus dysgalactiae	57	32	444
263	4	4411	3686	gi|1185002	dihydrodipicolinate reductase [Pseudomonas syringae pv. tabaci]	57	42	726
276	1	494	255	gi|396380	No definition line found [Escherichia coli]	57	40	240
283	2	335	1324	gi|773349	BirA protein [Bacillus subtilis]	57	32	990
297	1	469	236	gi|1334820	reading frame V [Cauliflower mosaic virus]	57	46	234
342	3	1993	2805	gi|1204431	hypothetical protein (SP:P13644) [Haemophilus influenzae]	57	35	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
375	6	3340	3741	gi|385177	cell division protein [Bacillus subtilis]	57	26	402
433	6	3286	4011	gi|1524117	alpha-acetolactate decarboxylase [Lactococcus lactis]	57	40	726
470	3	903	1145	gi|R04819	protein serine/threonine kinase [Toxoplasma gondii]	57	30	243
487	5	1391	1723	gi|507323	ORF1 [Bacillus stearothermophilus]	57	28	333
498	1	274	852	gi|1334549	NADH-ubiquinone oxidoreductase subunit 4L [Podospora anserina]	57	34*	579
503	1	343	173	gi|1502283	organic cation transporter OCT2 [Rattus norvegicus]	57	30	171
505	2	1619	1284	gi|1466884	B1496_C2_194 [Mycobacterium leprae]	57	40	336
519	2	1182	2549	gi|1303707	YrkH [Bacillus subtilis]	57	34	1368
522	2	3234	1945	gi|1064809	homologous to sp:MTRA_ECOLI [Bacillus subtilis]	57	36	1290
538	2	909	1415	gi|153179	phosphorinothyrcin n-acetyltransferase (Streptomyces coelicolor) pir|JH0246|JH0246 phosphinothricin N-acetyltransferase (EC 2.3.1.-) Streptomyces coelicolor	57	40	507
547	1	968	486	gi|467340	unknown [Bacillus subtilis]	57	50	483
559	1	1062	532	sp|P20692|TYRA_	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).	57	41	531
620	2	757	572	gi|1107894	unknown [Schizosaccharomyces pombe]	57	38	186
622	2	1600	1130	gi|173028	thiorodoxin II [Saccharomyces cerevisiae]	57	39	471
625	2	362	1114	gi|1262366	hypothetical protein [Mycobacterium leprae]	57	34	751
680	1	1	204	gi|143544	RNA polymerase sigma-30 factor [Bacillus subtilis] pir|A38625|A38625 transcription initiation factor sigma H - acillus subtilis	57	30	204
690	1	3	629	gi|466520	pocR [Salmonella typhimurium]	57	29	627
696	1	2	433	gi|413972	ipa-48r gene product [Bacillus subtilis]	57	33	432
704	1	36	638	gi|1499931	M. jannaschii predicted coding region MJ1083 [Methanococcus jannaschii]	57	36	603
732	1	2316	1621	gi|1418999	orf4 [Lactobacillus sake]	57	37	696
746	1	451	227	gi|392973	Rab3 [Aplysia californica]	57	42	225
757	1	20	466	gi|143979	L.curvatus small cryptic plasmid gene for rep protein [Lactobacillus rvatus]	57	45	447
862	1	2	295	gi|1303827	Yqfi [Bacillus subtilis]	57	21	294
1049	1	907	455	gi|1510108	ORF-1 [Agrobacterium tumefaciens]	57	35	453
1117	1	1387	695	gi|896286	NN2 terminus uncertain [Leishmania tarentolae]	57	28	693

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1136	1	2	322	gi|1301853	YqgF [Bacillus subtilis]	57	38	321
1144	2	1033	611	gi|310083	voltage-activated calcium channel alpha-1 subunit (Rattus orvegicus)	57	46	423
1172	1	1472	738	gi|1511146	M. jannaschii predicted coding region MJ1143 (Methanococcus jannaschii)	57	28	735
1500	2	746	558	gi|142780	putative membrane protein; putative (Bacillus subtilis)	57	35	189
1676	1	659	399	gi|313777	uracil permease (Escherichia coli)	57	31	261
2481	1	2	400	gi|1237015	ORF4 (Bacillus subtilis)	57	23	399
3099	1	3	230	gi|1204540	isochorismate synthase (Haemophilus influenzae)	57	39	228
3122	1	360	181	gi|882472	ORF_o464 (Escherichia coli)	57	40	180
3560	1	2	361	gi|153490	tetracenomycin C resistance and export protein (Streptomyces lauceacene)	57	37	360
3850	1	856	434	gi|1155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	57	40	423
3931	1	704	354	gi|1413953	ipa-29d gene product (Bacillus subtilis)	57	36	351
3992	1	1	384	gi|151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	57	39	384
4065	1	793	398	pir|JV0037|RDEC	nitrate reductase (EC 1.7.99.4) alpha chain - Escherichia coli	57	31	396
4100	1	596	300	gi|1086633	T06C10.5 gene product (Caenorhabditis elegans)	57	47	297
4163	1	571	287	gi|21512	patatin (Solanum tuberosum)	57	50	285
4267	2	631	335	gi|1000365	SpoIIIAG (Bacillus subtilis)	57	38	297
4358	1	3	302	gi|298012	EF (Streptococcus suis)	57	32	300
4389	2	108	290	gi|405894	1-phosphofructokinase (Escherichia coli)	57	37	183
4399	1	2	232	gi|1483603	Pristinamycin I synthase I (Streptomyces pristinaespiralis)	57	35	231
4481	1	572	288	gi|405879	yeiH (Escherichia coli)	57	44	285
4466	1	512	258	gi|1515938	glutamate synthase (ferredoxin) (Symechocystis sp.) pir|S46957|S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	57	42	255
4510	1	461	242	gi|1205301	leukotoxin secration ATP-binding protein (Haemophilus influenzae)	57	38	240
4617	1	468	256	gi|1511222	restriction modification enzyme, subunit M (Methanococcus jannaschii)	57	35	213
4	11	12201	11524	gi|149204	histidine utilization repressor G (Klebsiella aerogenes) pir|A36730|A36730 hutG protein - Klebsiella pneumoniae (fragment) sp|P19452|HUTG_KLEAE FORMIMINOGLUTAMASE (EC 3.5.3.8) FORMIMINOGLUTAMATE HYDROLASE) (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT).	56	31	678

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
22	8	4248	5177	gi|1322222	RACH1 [Homo sapiens]	56	33	930
38	28	21179	22264	gi|1480705	lipoate-protein ligase [Mycoplasma capricolum]	56	34	1086
44	3	1861	2421	gi|490320	Y gene product (unidentified)	56	31	561
44	15	10103	10606	gi|1205099	hypothetical protein (GB:L19201_1) [Haemophilus influenzae]	56	39	504
50	6	4820	5161	gi|209931	fiber protein [Human adenovirus type 5]	56	48'	342
53	4	2076	2972	gi|623476	transcriptional activator [Providencia stuartii] sp|P43463|AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP.	56	30	897
67	6	5656	6594	gi|466613	nikB [Escherichia coli]	56	32	939
89	3	2364	1810	gi|482922	protein with homology to pai repressor of B.subtilis [Lactobacillus elbrueckii]	56	39	555
96	1	203	913	gi|145594	cAMP receptor protein (crp) [Escherichia coli]	56	35	711
109	21	18250	17846	gi|1204367	hypothetical protein (GB:U14001_278) [Haemophilus influenzae]	56	27	405
112	8	5611	6678	gi|155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	56	40	1068
131	3	6404	5100	gi|619724	MgtE [Bacillus firmus]	56	30	1305
138	2	65	232	gi|413948	ipa-24d gene product [Bacillus subtilis]	56	31	168
138	4	823	1521	gi|580868	ipa-22r gene product [Bacillus subtilis]	56	31	699
146	2	740	447	gi|1046009	M. genitalium predicted coding region MG309 [Mycoplasma genitalium]	56	37	294
149	2	1639	1067	gi|945180	terminase small subunit [Bacteriophage LL-II]	56	35	573
163	1	2	223	gi|143947	glutamine synthetase [Bacteroides fragilis]	56	30	222
166	5	6745	6449	gi|405792	ORF154 [Pseudomonas putida]	56	26	297
187	1	31	393	gi|311237	H(+)-transporting ATP synthase [Zea mays]	56	30	363
190	1	2	373	gi|1109686	ProX [Bacillus subtilis]	56	35	372
191	8	11538	9943	gi|581070	acyl coenzyme A synthetase [Escherichia coli]	56	35	1596
195	1	1291	647	gi|1510242	collagenase [Methanococcus jannaschii]	56	34	645
230	3	2323	2072	gi|40363	heat shock protein [Clostridium acetobutylicum]	56	39	252
238	5	3383	3775	gi|1477533	sarA [Staphylococcus aureus]	56	31	393
270	2	813	1712	gi|765073	autolysin [Staphylococcus aureus]	56	41	900

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
290	1	3221	1632	gi|547313	orf3 [Haemophilus influenzae]	56	34	1590
297	5	1140	1373	gi|1511556	M. jannaschii predicted coding region MJ1561 [Methanococcus jannaschii]	56	40	234
321	2	2947	1799	gi|1001801	hypothetical protein [Synechocystis sp.]	56	31	1149
359	2	1279	641	gi|46336	nolI gene product [Rhizobium meliloti]	56	26	639
371	2	360	1823	gi|145304	L-ribulokinase [Escherichia coli]	56	39	1464
391	4	1762	2409	gi|1001634	hypothetical protein [Synechocystis sp.]	56	34	648
402	1	380	192	gi|1438904	5-HT4L receptor [Homo sapiens]	56	48	189
416	4	2480	2109	gi|1408486	HS74A gene product [Bacillus subtilis]	56	31	372
424	3	1756	2334	gi|142471	acetolactate decarboxylase [Bacillus subtilis]	56	32	579
457	1	1907	1017	gi|1205194	formamidopyrimidine-DNA glycosylase [Haemophilus influenzae]	56	36	891
458	2	2423	1812	gi|15466	terminase [Bacteriophage SPP1]	56	37	612
504	2	2152	1283	gi|1142681	Lpp38 [Pasteurella haemolytica]	56	38	870
511	1	1	1284	gi|217049	brnQ protein [Salmonella typhimurium]	56	37	1284
604	3	1099	1701	gi|467109	rim; 30S Ribosomal protein S18 alanine acetyltransferase; 229_C1_170 [Mycobacterium leprae]	56	43	603
660	5	3547	3774	gi|1229106	2K930.1 [Caenorhabditis elegans]	56	30	228
707	1	35	400	gi|153929	NADPH-sulfite reductase flavoprotein component [Salmonella yphimurium]	56	38	366
709	2	1385	1095	gi|1510801	hydrogenase accessory protein [Methanococcus jannaschii]	56	38	291
718	1	1	495	gi|411948	ipa-24d gene product [Bacillus subtilis]	56	35	495
743	1	87	677	gi|928836	repressor protein [Lactococcus lactis phage BK5-T]	56	35	591
790	1	776	399	gi|1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	56	37	378
795	1	3	407	gi|1205382	cell division protein [Haemophilus influenzae]	56	34	405
813	1	19	930	gi|1222161	permease [Haemophilus influenzae]	56	28	912
855	1	3	515	gi|1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	56	33	513
968	1	2	466	gi|547313	orf3 [Haemophilus influenzae]	56	37	465
973	2	1049	732	gi|886022	MexR [Pseudomonas aeruginosa]	56	71	318
1203	1	5	223	gi|184251	HMG-1 [Homo sapiens]	56	34	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1976	1	452	237	gi|9906	lysine-rich aspartic acid-rich protein [Plasmodium chabaudi] r|S22185|S22183 lysine/aspartic acid-rich protein - Plasmodium baudi	56	33	216
2161	1	2	400	gi|1237015	ORF4 [Bacillus subtilis]	56	27	399
2958	1	362	183	gi|466685	No definition line found [Escherichia coli]	56	26	180
2979	1	421	212	gi|1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	56	40	210
2994	2	526	326	gi|836646	phosphoribosylformimino-praic ketoisomerase [Rhodobacter phaeroides]	56	29	201
3026	1	177	328	gi|143306	penicillin V amidase [Bacillus sphaericus]	56	30	150
3189	1	289	146	gi|1166604	Similar to aldehyde dehydrogenase [Caenorhabditis elegans]	56	37	144
3770	1	63	401	gi|1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	56	43	339
4054	2	720	361	gi|1205355	Na+/H+ antiporter [Haemophilus influenzae]	56	31	360
4145	1	1	324	gi|726095	long-chain acyl-CoA dehydrogenase [Mus musculus]	56	36	324
4200	1	505	254	gi|155588	glucose-fructose oxidoreductase [Zymomonas mobilis] pir|A42289|A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) recursor - Zymomonas mobilis	56	40	252
4273	1	675	355	gi|308861	GTG start codon [Lactococcus lactis]	56	33	321
1	3	4095	3436	gi|5341	Putative orf YCLX8c, len:192 [Saccharomyces cerevisiae] r|S53591|S53591 hypothetical protein - yeast (Saccharomyces evisiae)	55	25	660
11	12	9377	8505	gi|216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	873
12	4	5133	4534	gi|467337	unknown [Bacillus subtilis]	55	26	600
19	5	5404	5844	gi|1001719	hypothetical protein [Synechocystis sp.]	55	25	441
23	13	14087	12339	gi|474190	lucA gene product [Escherichia coli]	55	30	1749
32	7	5368	6888	gi|1140096	unknown [Mycobacterium tuberculosis]	55	37	1521
34	3	2569	1808	gi|1303968	YqjQ [Bacillus subtilis]	55	39	762
34	5	3960	3412	gi|1303962	YqjK [Bacillus subtilis]	55	33	549
36	1	1291	647	gi|606045	ORF_o118 [Escherichia coli]	55	27	645
36	6	6320	5243	gi|1001341	hypothetical protein [Synechocystis sp.]	55	31	978
47	3	3054	3821	gi|1001819	hypothetical protein [Synechocystis sp.]	55	21	768
49	1	2065	1127	gi|1403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir|S37251|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	55	36	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% aim	% ident	length (nt)
67	11	8966	9565	gi|151053	norA1199 protein (Staphylococcus aureus)	55	23	600
75	3	881	1273	gi|51698	L-histidinol: NAD+ oxidoreductase (EC 1.1.1.23) (aa 1-434) scherichia coli)	55	33	393
82	9	15387	14194	gi|136221	carboxypeptidase (Sulfolobus solfataricus)	55	35	1194
87	4	3517	4917	gi|1068812	function unknown (Bacillus subtilis)	55	26	1401
88	2	1172	1636	gi|882463	protein-N(pi)-phosphohistidine-sugar phosphotransferase (Escherichia coli)	55	35	465
92	1	127	516	gi|1377832	unknown (Bacillus subtilis)	55	36	390
100	2	836	2035	gi|1170274	zeaxanthin epoxidase (Nicotiana plumbaginifolia)	55	36	1200
100	5	5137	4658	gi|396660	unknown open reading frame (Buchnera aphidicola)	55	29	480
108	3	4266	2986	gi|1499866	M. jannaschii predicted coding region MJ1024 (Methanococcus jannaschii)	55	31	1281
114	3	2616	1834	gi|1511367	formate dehydrogenase, alpha subunit (Methanococcus jannaschii)	55	29	783
144	3	1805	1476	gi|1100787	unknown (Saccharomyces cerevisiae)	55	35	330
165	5	6212	5508	gi|1045884	M. genitalium predicted coding region MG199 (Mycoplasma genitalium)	55	27	705
189	5	2205	2576	gi|142569	ATP synthase a subunit (Bacillus firmus)	55	35	372
191	6	9136	6857	gi|559411	B0272.3 (Caenorhabditis elegans)	55	39	2280
194	2	364	636	gi|1145768	K7 kinesin-like protein (Dictyostelium discoideum)	55	34	273
209	4	1335	1676	gi|473357	thi4 gene product (Schizosaccharomyces pombe)	55	35	342
211	2	1693	1145	gi|410130	ORFX6 (Bacillus subtilis)	55	37	549
213	2	644	1372	gi|633692	TraA (Yersinia enterocolitica)	55	28	729
214	7	4144	5481	gi|1001793	hypothetical protein (Synechocystis sp.)	55	30	1338
221	7	11473	9197	gi|466520	pocR (Salmonella typhimurium)	55	32	2277
233	8	5908	4817	gi|1237063	unknown (Mycobacterium tuberculosis)	55	18	1092
236	4	1375	2340	gi|1146199	putative (Bacillus subtilis)	55	32	966
243	2	380	1885	gi|459907	mercuric reductase (Plasmid pI258)	55	29	1506
258	1	786	394	gi|455006	orf6 (Rhodococcus fascians)	55	36	393
281	1	126	938	gi|1408493	homologous to SwissProt:YIDA_ECOLI hypothetical protein (Bacillus subtilis)	55	35	813
316	3	1323	2102	gi|1486447	luxA homologue (Rhizobium sp.)	55	30	780
326	5	2968	2744	gi|1296824	proline iminopeptidase (Lactobacillus helveticus)	55	36	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
351	2	2322	1429	gi|1204820	hydrogen peroxide-inducible activator (Haemophilus influenzae)	55	28	894
353	4	2197	2412	gi|1272475	chitin synthase (Emericella nidulans)	55	50	216
380	1	14	379	gi|142554	ATP synthase 1 subunit (Bacillus megaterium)	55	37	366
383	1	462	232	gi|349272	ferrichrome-binding protein (Bacillus subtilis)	55	36	231
386	1	3	938	gi|1510251	DNA helicase, putative (Methanococcus jannaschii)	55	30	936
410	2	1208	1891	gi|1205144	multidrug resistance protein (Haemophilus influenzae)	55	27	684
483	2	411	833	gi|413934	ipa-10r gene product (Bacillus subtilis)	55	26	423
529	3	1777	1433	gi|606150	ORF_f309 (Escherichia coli)	55	33	345
555	1	1088	585	gi|143407	para-aminobenzoic acid synthase, component I (pab) (Bacillus subtilis)	55	28	504
565	1	402	202	gi|1223961	CDP-tyvelose epimerase (Yersinia pseudotuberculosis)	55	41	201
582	1	751	452	gi|1256643	20.2% identity with NADH dehydrogenase of the Leishmania major mitochondrion; putative (Bacillus subtilis)	55	36	300
645	5	2260	2057	gi|210824	fusion protein F [Bovine respiratory syncytial virus] pir|JQ1481|VGNZBA (fusion glycoprotein precursor - bovine respiratory syncytial virus (strain A51908)	55	25	204
672	2	957	2216	gi|1511333	M. jannaschii predicted coding region MJ1322 (Methanococcus jannaschii)	55	36	1260
730	1	955	479	gi|537007	ORF_f379 (Escherichia coli)	55	30	477
737	1	1859	945	gi|536963	CG site No. 18166 (Escherichia coli)	55	30	915
742	2	228	572	gi|304160	product unknown (Bacillus subtilis)	55	18	345
817	2	1211	903	gi|1136289	histidine kinase A (Dictyostelium discoideum)	55	29	309
819	1	582	355	gi|558073	polymorphic antigen (Plasmodium falciparum)	55	22	228
832	2	1152	724	gi|40367	ORFC (Clostridium acetobutylicum)	55	32	429
840	1	769	386	gi|1205875	pseudouridylate synthase I (Haemophilus influenzae)	55	39	384
1021	1	23	529	gi|48563	beta-lactamase (Yersinia enterocolitica)	55	38	507
1026	1	60	335	gi|47804	Opp C (AA1-301) (Salmonella typhimurium)	55	26	276
1525	1	1	282	gi|1477533	sarA (Staphylococcus aureus)	55	29	282
1814	2	224	985	gi|1046078	M. genitalium predicted coding region MG369 (Mycoplasma genitalium)	55	38	762
2254	1	427	254	gi|413968	ipa-44d gene product (Bacillus subtilis)	55	30	174

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
1695	1	686	345	gi|216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	342
1721	1	1	312	gi|42029	ORF1 gene product [Escherichia coli]	55	31	312
1799	1	3	272	gi|42029	ORF1 gene product [Escherichia coli]	55	36	270
3889	1	22	423	gi|1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	45	402
3916	1	2	385	gi|529754	spec [Streptococcus pyogenes]	55	38	384
3945	1	4	198	gi|476252	phase 1 flagellin [Salmonella enterica]	55	36	195
4074	1	488	246	gi|42029	ORF1 gene product [Escherichia coli]	55	38	243
4184	1	2	343	gi|524267	unknown [Mycobacterium tuberculosis]	55	28	342
4284	1	14	208	gi|1100774	ferredoxin-dependent glutamate synthase (Synechocystis sp.)	55	36	195
4457	2	644	378	gi|180189	cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi|182737 cerebellar degeneration-associated protein [Homo sapiens] pir|A29770|A29770 cerebellar degeneration-related protein - human	55	38	267
4514	1	2	244	gi|216773	haloacetate dehalogenase H-1 [Moraxella sp.]	55	32	243
4599	1	432	217	gi|1129145	acetyl-CoA C-acyltransferase [Mangifera indica]	55	42	216
4606	1	416	210	gi|386120	myosin alpha heavy chain (S2 subfragment) [rabbits, masseter, eptide Partial, 234 aa]	55	27	207
5	8	5348	4932	gi|556069	ORF YDL047c [Saccharomyces cerevisiae]	54	27	417
12	7	7166	6165	gi|1205504	homoserine acetyltransferase [Haemophilus influenzae]	54	30	1002
23	16	17086	15326	gi|474192	iucC gene product [Escherichia coli]	54	31	1761
35	1	2	979	gi|48054	small subunit of soluble hydrogenase (AA 1-384) [Synechococcus sp.] lr|S06919|HQYCSS soluble hydrogenase (EC 1.12.--) small chain - nchococcus sp. (PCC 6716)	54	36	978
37	11	9417	8667	gi|337207	ORF_f277 [Escherichia coli]	54	38	771
37	12	8165	8332	gi|1160967	palmitoyl-protein thioesterase [Homo sapiens]	54	37	168
46	15	13025	13804	gi|438473	protein is hydrophobic, with homology to E. coli ProW; putative Bacillus subtilis)	54	28	780
56	2	203	726	gi|1256139	YbbJ [Bacillus subtilis]	54	34	534
57	13	11117	110179	gi|1151248	inosine-uridine preferring nucleoside hydrolase [Crithidia fasciculata]	54	32	939
66	2	516	1133	gi|1335781	Cap [Drosophila melanogaster]	54	29	618
70	10	8116	8646	gi|1399823	PhoE [Rhizobium meliloti]	54	31	531

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
70	15	12556	11801	sp|P02983|TCR_S	TETRACYCLINE RESISTANCE PROTEIN.	54	29	756
87	5	4915	5706	gi|1064811	[function unknown [Bacillus subtilis]	54	33	792
92	4	3005	2289	gi|1205366	oligopeptide transport ATP-binding protein [Haemophilus influenzae]	54	33	717
103	2	2556	1556	gi|710495	protein kinase [Bacillus brevis]	54	33	1041
105	2	3585	2095	gi|143727	putative [Bacillus subtilis]	54	30'	1491
112	4	2317	2712	gi|153724	MalC [Streptococcus pneumoniae]	54	41	396
127	2	1720	2493	gi|144297	acetyl esterase (XynC) [Caldocellum saccharolyticum] pir|B37202|B37202 acetylesterase (EC 3.1.1.6) (XynC) - Caldocellum saccharolyticum	54	34	774
13A	5	1600	3306	gi|42471	pyruvate oxidase [Escherichia coli]	54	36	1707
152	2	525	1172	gi|1377834	unknown [Bacillus subtilis]	54	23	648
161	9	4831	5469	gi|903305	ORF73 [Bacillus subtilis]	54	28	639
161	13	6694	7251	gi|1511039	phosphate transport system regulatory protein [Methanococcus jannaschii]	54	32	558
164	6	1263	4543	gi|1204976	prolyl-tRNA synthetase [Haemophilus influenzae]	54	34	1281
164	20	21602	22243	gi|143582	SpoIIEA protein [Bacillus subtilis]	54	32	642
171	6	5683	4250	gi|1436965	[malA] gene products [Bacillus stearothermophilus] pir|S43914|S43914 hypothetical protein 1 - Bacillus tearothermophilus	54	37	1434
206	1A	19208	19720	gi|1240016	R09E10.3 [Caenorhabditis elegans]	54	38	513
218	2	1090	1905	gi|467378	unknown [Bacillus subtilis]	54	26	816
220	1	1322	663	gi|1353761	myosin II heavy chain [Naegleria fowleri]	54	22	660
220	13	12655	11059	pir|S00485|S004	gene 11-1 protein precursor - Plasmodium falciparum (fragment)	54	35	405
221	3	2030	3709	gi|1303813	YqeW [Bacillus subtilis]	54	34	1680
272	7	5055	4219	gi|62964	arylamine N-acetyltransferase (AA 1-290) [Gallus gallus] ir|S06652|XYCHY3 arylamine N-acetyltransferase (EC 2.3.1.5) (clone NAT-3) - chicken	54	33	837
316	7	4141	4701	gi|682769	mccE gene product [Escherichia coli]	54	31	561
316	10	6994	8742	gi|413951	ipa-27d gene product [Bacillus subtilis]	54	28	1749
338	3	3377	2214	gi|490328	LORF F [unidentified]	54	28	1164
341	4	3201	3614	gi|171959	myosin-like protein [Saccharomyces cerevisiae]	54	25	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	sim	% ident.	Length (nt)
346	1	1820	912	gi|396400	similar to eukaryotic Na+/H+ exchangers (Escherichia coli) sp|P12703|YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS INTERGENIC REGION (O549).	54	34	909
348	2	623	1351	gi|537109	ORF_f343a (Escherichia coli)	54	34	729
378	2	1007	1942	sp|P02983|TCR_S	TETRACYCLINE RESISTANCE PROTEIN.	54	3?	936
408	6	4351	5301	gi|474190	iucA gene product (Escherichia coli)	54	29*	951
444	9	7934	8854	gi|216267	ORF2 (Bacillus megaterium)	54	32	921
463	2	2717	2229	gi|104160	product unknown (Bacillus subtilis)	54	50	489
502	2	1696	1133	gi|1205015	hypothetical protein (SP:P10120) (Haemophilus influenzae)	54	38	564
505	6	6262	5357	gi|1500558	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (Methanococcus jannaschii)	54	61	906
550	1	2736	1522	gi|40100	rodc (tagI) polypeptide (AA 1-746) [Bacillus subtilis] tr|S06049|S06049 rodc protein - Bacillus subtilis p|P13485|TAGP_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN F.	54	35	1215
551	5	3305	4279	gi|950197	unknown (Corynebacterium glutamicum)	54	34	975
558	2	1156	958	gi|485090	No definition line found (Caenorhabditis elegans)	54	32	399
580	1	91	936	gi|331906	(fused envelope glycoprotein precursor (Friend spleen focus-forming [rus]	54	45	846
603	3	554	757	gi|1323423	ORF YGR234w (Saccharomyces cerevisiae)	54	26	204
617	1	25	249	gi|219959	ornithine transcarbamylase (Homo sapiens)	54	40	225
622	3	1097	1480	gi|1303873	Yqg2 (Bacillus subtilis)	54	25	384
623	1	3	404	gi|1063250	low homology to P20 protein of Bacillus lichiniformis and bleomycin acetyltransferase of Streptomyces verticillus (Bacillus subtilis)	54	45	402
649	1	1547	1011	gi|552446	NADH dehydrogenase subunit 4 (Apis mellifera ligustica) pir|S52968|S52968 NADH dehydrogenase chain 4 - honeybee itochondrion (SGC4)	54	30	537
725	2	686	1441	gi|987096	sensory protein kinase (Streptomyces hygroscopicus)	54	26	756
956	1	1	249	pir|S30782|S307	integrin homolog - yeast (Saccharomyces cerevisiae)	54	24	249
978	2	1137	859	gi|1301994	ORF YNL091w (Saccharomyces cerevisiae)	54	33	279
1114	1	3	281	gi|1001108	hypothetical protein (Synechocystis sp.)	54	33	279
2450	1	1	228	gi|1045057	ch-TOG (Homo sapiens)	54	32	228
2934	1	1	387	gi|580870	ipa-17d gpxA gene product (Bacillus subtilis)	54	36	387
2970	1	499	251	sp|P1734P|VECE	HYPOTHETICAL PROTEIN IN ASPS 5'REGION (FRAGMENT).	54	42	249

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3002	1	1	309	gi|34027	[Tma protein (Lactococcus lactis)]	54	33	309
3561	1	9	464	gi|151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir|A44756|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	54	35	456
3572	1	72	401	gi|450688	hadM gene of Ecoprri gene product [Escherichia coli] pir|S38437|S38437 hadM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	54	36	330
3829	1	798	400	gi|1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	54	29	399
3909	1	1	273	gi|29865	(CENP-E (Homo sapiens)	54	30	273
3921	1	1	209	pir|S24325|S243	glucan 1.4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa	54	34	207
4438	1	566	285	gi|1196657	unknown protein [Mycoplasma pneumoniae]	54	30	282
4459	1	3	272	gi|1046081	hypothetical protein (GB:D26185_10) (Mycoplasma genitalium)	54	38	270
4564	1	3	221	gi|216267	ORF2 [Bacillus megaterium]	54	38	219
23	12	12538	10685	gi|474192	lucC gene product [Escherichia coli]	53	35	1854
23	14	14841	13579	gi|42029	ORF1 gene product [Escherichia coli]	53	32	1263
24	3	4440	3940	gi|1369947	c2 gene product [Bacteriophage B1]	53	36	501
26	4	3818	4618	gi|1486287	unknown [Bacillus subtilis]	53	37	801
38	6	2856	3998	gi|405880	yell [Escherichia coli]	53	40	1143
38	10	9380	7806	gi|1399954	thyroid sodium/iodide symporter NIS [Rattus norvegicus]	53	29	1575
56	10	12324	12100	pir|A54592|A545	110k actin filam.. associated protein - chicken	53	32	225
57	6	5047	4583	pir|A00341|DEZP	alcohol dehydrogenase (EC 1.1.1.1) - fission yeast (Schizosaccharomyces pombe)	53	39	465
57	12	10515	8932	gi|480429	putative transcriptional regulator [Bacillus stearothermophilus]	53	30	1584
67	12	9496	10218	gi|1511555	quinolone resistance norA protein protein [Methanococcus jannaschii]	53	31	723
69	3	3125	2382	gi|1687017	arabinogalactan-protein, AGP [Nicotiana alata, cell-suspension culture (filtrate, Peptide, 461 aa]	53	30	744
79	1	3	1031	gi|1523802	glucanase [Anabaena variabilis]	53	32	1029
80	1	673	338	gi|452428	ATPase 3 [Plasmodium falciparum]	53	36	336
88	4	1910	2524	gi|137034	ORP_o488 [Escherichia coli]	53	25	615
88	5	2467	3282	gi|537034	ORP_o488 [Escherichia coli]	53	29	816

TABLE 2

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
92	8	5870	5505	gi|399598	amphotropic murine retrovirus receptor [Rattus norvegicus]	53	33	366
94	5	4417	3239	gi|173038	tropomyosin (TPM1) [Saccharomyces cerevisiae]	53	25	1179
99	5	4207	3433	sp|P28246|BCR_E	BICYCLOMYCIN RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN).	53	30	1227
120	3	1639	2262	gi|576655	ORF1 [Vibrio anguillarum]	53	35	624
120	11	7257	8897	gi|1524397	glycine betaine transporter OpuD [Bacillus subtilis]	53	33'	1641
127	6	6893	5685	gi|1256630	putative [Bacillus subtilis]	53	32	1209
147	2	255	557	gi|581648	epi8 gene product [staphylococcus epidermidis]	53	34	303
158	4	4705	4256	gi|151004	mucoidy regulatory protein AlgR [Pseudomonas aeruginosa] pir|A32802|A32802 regulatory protein algR - Pseudomonas aeruginosa sp|P26275|ALGR_PSEAE POSITIVE ALGINATE BIOSYNTHESIS REGULATORY PROTEIN.	53	32	450
171	7	5717	5421	gi|510669	hypothetical protein (GP:D64044_18) [Methanococcus jannaschii]	53	34	297
191	9	13087	11483	gi|298085	acetoacetate decarboxylase [Clostridium acetobutylicum] pir|B49346|B49346 butyrate-acetoacetate CoA-transferase (EC .8.3.9) small chain - Clostridium acetobutylicum sp|P31752|CTFA_CLOAB BUTYRATB-ACETOACETATE COA-TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)	53	31	1605
203	5	3763	4326	gi|143456	rpoE protein (ttg start codon) [Bacillus subtilis]	53	29	564
206	17	18204	18971	gi|304136	acetylglutamate kinase [Bacillus stearothermophilus] sp|007905|ARGB_BACST ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG INASE) (AGK) (N-ACETYL-L-GLUTAMATE 5-PHOSPHOTRANSFERASE].	53	36	768
212	10	4021	4221	gi|9878	protein kinase [Plasmodium falciparum]	53	28	201
231	2	1580	1350	gi|537506	paramyosin [Dirofilaria immitis]	53	34	231
272	6	2719	3249	pir|A33141|A331	hypothetical protein (gtfD 3' region) - Streptococcus mutans	53	34	531
308	3	927	2576	gi|606292	ORF_o696 [Escherichia coli]	53	34	1650
320	7	5645	5884	gi|160596	RNA polymerase III largest subunit [Plasmodium falciparum] sp|P27625|RPC1_PLAFA DNA-DIRECTED RNA POLYMERASE III LARGEST UBUNIT (EC 2.7.7.6).	53	33	240
327	1	218	901	gi|854601	unknown [Schizosaccharomyces pombe]	53	31	684
341	2	212	2500	gi|613732	ORF1 [Campylobacter jejuni]	53	31	2289
351	1	763	383	sp|P31675|YABM_	HYPOTHETICAL 42.7 KD PROTEIN IN TBPA-LEUD INTERGENIC REGION (ORF104).	53	32	381
433	7	5087	4731	gi|1001961	MHC class II analog [Staphylococcus aureus]	53	30	357
454	2	1240	980	pir|A60128|A603	40K cell wall protein precursor (sr 5' region) - Streptococcus mutans (strain OMZ175, serotype f)	53	27	261

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	AmIgh gene name	% sim	% ident	length (nt)
470	4	1123	1761	gi|516826	rat GCP360 (Rattus rattus)	53	30	639
483	1	432	217	gi|1480429	putative transcriptional regulator [Bacillus stearothermophilus]	53	23	216
544	1	516	1259	gi|146587	ORF 1 (AA 1 - 121) (1 is 2nd base in codon) [Staphylococcus aureus] ir|S15765|S15765 hypothetical protein 1 (hib 5' region) - aphylococcus aureus (fragment)	53	38	744
558	10	3957	3754	gi|15140	res gene [Bacteriophage P1]	53	32'	204
603	2	339	620	gi|507738	Hmp [Vibrio parahaemolyticus]	53	26	282
693	1	1669	941	gi|151123	toxic shock syndrome toxin-1 precursor [Staphylococcus aureus] pir|A24606|XCSAS1 toxic shock syndrome toxin-1 precursor - taphylococcus aureus	53	38	729
766	1	2	673	gi|687600	orfA2; orfA2 forms an operon with orfA1 [Listeria monocytogenes]	53	43	672
781	1	667	335	gi|1204551	pilin biogenesis protein [Haemophilus influenzae]	53	26	333
801	1	3	545	gi|279400	SapA protein [Escherichia coli]	53	25	543
803	1	2	910	gi|695278	lipase-like enzyme [Alcaligenes eutrophus]	53	30	909
872	1	1177	590	gi|298032	EP [Streptococcus suis]	53	30	588
910	1	2	184	gi|1044936	unknown [Schizosaccharomyces pombe]	53	29	183
943	1	794	399	gi|190508	similar to unidentified ORF near 47 minutes [Escherichia coli] sp|P31436|YICK_ECOLI HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA INTERGENIC REGION.	53	30	396
988	1	1004	504	gi|142441	ORF 3; putative [Bacillus subtilis]	53	28	501
1064	1	3	434	gi|305080	myosin heavy chain [Entamoeba histolytica]	53	26	432
1366	2	3	452	gi|308852	transmembrane protein [Lactococcus lactis]	53	33	450
1758	1	792	397	gi|1001774	hypothetical protein [Synechocystis sp.]	53	30	396
1897	1	1	447	gi|1303949	Yqix [Bacillus subtilis]	53	27	447
2381	1	798	400	gi|1116243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	53	37	399
3537	1	1	327	gi|450688	hsdM gene of Ecoprrl gene product (Escherichia coli) pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	53	35	327
3747	2	137	397	gi|1477486	transposase [Burkholderia cepacia]	53	53	261
11	5	3049	3441	gi|868224	No definition line found [Caenorhabditis elegans]	52	33	393

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
15	5	2205	2369	gi|215966	G41 protein (gtg start codon) (Bacteriophage T4)	52	34	165
19	3	2429	3808	gi|1205379	UDP-murnac-pentapeptide synthetase (Haemophilus influenzae)	52	31	1380
24	1	6920	3462	gi|579124	predicted 86.4kd protein; 52kd observed (Mycobacteriophage 15) pir|S30971|S30971 gene 26 protein - Mycobacterium phage L5 sp|Q05233|VG26_BPML5 MINOR TAIL PROTEIN QP26. (SUB 2-837)	52	32	3459
37	5	3015	3935	gi|1500543	P115 protein (Methanococcus jannaschii)	52	25	921
38	13	8795	9703	gi|46851	glucose kinase (Streptomyces coelicolor)	52	29	909
44	16	10617	11066	gi|42012	moaE gene product (Escherichia coli)	52	36	450
46	1	3	521	gi|1040957	NADH dehydrogenase subunit 6 (Anopheles trinkae)	52	25	519
51	10	5531	6280	gi|388269	traC (Plasmid pADl)	52	32	750
56	5	3968	2826	gi|181949	endothelial differentiation protein (edg-1) (Homo sapiens) pir|A35300|A35300 G protein-coupled receptor edg-1 - human sp|P21453|EDG1_HUMAN PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-1.	52	23	1143
57	5	4850	4173	gi|304153	sorbitol dehydrogenase (Bacillus subtilis)	52	27	678
62	5	3364	2870	gi|1072399	phaE gene product (Rhizobium meliloti)	52	25	495
62	6	4445	3651	gi|46485	NADH dehydrogenase (Synechococcus PCC7942)	52	27	795
67	14	11355	12962	gi|1511365	glutamate synthase (NADPH), subunit alpha (Methanococcus jannaschii)	52	30	1608
67	21	16935	18158	gi|1204393	hypothetical protein (SP:P31122) (Haemophilus influenzae)	52	25	1224
70	4	2185	1997	gi|7227	cytoplasmic dynein heavy chain (Dictyostelium discoideum) r|A44357|A44357 dynein heavy chain, cytosolic - slime mold (Dictyostelium discoideum)	52	16	189
96	10	10005	10664	gi|1408485	B650 gene product (Bacillus subtilis)	52	26	660
103	5	3986	3351	gi|1009368	Respiratory nitrate reductase (Bacillus subtilis)	52	42	636
109	3	4102	3350	gi|699274	lmbE gene product (Mycobacterium leprae)	52	39	753
109	19	15732	17300	gi|1526981	amino acid permease YeeF like protein (Salmonella typhimurium)	52	30	1569
121	3	1412	981	gi|732931	unknown (Saccharomyces cerevisiae)	52	32	432
125	3	865	1680	gi|1296975	puT gene product (Porphyromonas gingivalis)	52	18	816
130	2	659	1807	gi|1256634	25.8% identity over 120 aa with the Synenococcus sp. MpeV protein; putative MpeV protein; putative	52	36	1149
149	1	1164	583	gi|1225943	PB6X terminase (Bacillus subtilis)	52	33	582
149	14	4687	4415	gi|1510368	M. jannaschii predicted coding region MJ0272 (Methanococcus jannaschii)	52	35	273

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
167	1	216	1001	gi|146025	cell division protein [Escherichia coli]	52	43	786
188	1	120	1256	gi|474915	orf 337; translated orf similarity to SW: BCR_ECOLI bicyclomycin esistance protein of Escherichia coli [Coxiella burnetii] pir|S44207|S44207 hypothetical protein 337 - Coxiella burnetii (SUB -338)	52	26	1137
195	9	9161	8760	gi|3028	mitochondrial outer membrane 72K protein [Neurospora crassa] r|A36682|A36682 72K mitochondrial outer membrane protein - rospora crassa	52	25	402
200	1	2065	2607	gi|142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	543
203	4	2776	3684	gi|1301698	BltD [Bacillus subtilis]	52	25	909
227	8	5250	5651	gi|305080	myosin heavy chain [Entamoeba histolytica]	52	24	402
242	1	21	1424	gi|1060877	EmrY [Escherichia coli]	52	32	1404
249	5	4526	4753	pir|C37222|C372	cytochrome P450 1A1. hepatic - dog (fragment)	52	23	228
255	1	2107	1055	gi|143290	penicillin-binding protein [Bacillus subtilis]	52	28	1053
276	7	3963	3664	gi|1001610	hypothetical protein [Synechocystis sp.]	52	30	300
276	8	4456	4055	gi|416235	orf L3 [Mycoplasma capricolum]	52	26	402
289	2	1856	1449	gi|150900	GTP phosphohydrolase [Proteus vulgaris]	52	34	408
325	1	1	279	gi|1204874	polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]	52	33	279
340	1	2017	1010	gi|1215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	52	33	1008
375	3	340	1878	gi|467446	similar to SpoVB [Bacillus subtilis]	52	28	1539
424	4	4104	3262	gi|1478239	unknown [Mycobacterium tuberculosis]	52	34	843
430	1	3	575	pir|A42606|A426	orfA 5' to orf405 - Saccharopolyspora erythraea (fragment)	52	28	573
444	4	4728	3712	gi|1408494	homologous to penicillin acylase [Bacillus subtilis]	52	31	1017
465	1	1802	901	gi|143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir|A27650|A27650 regulatory protein phoR - Bacillus subtilis sp|P21545|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-)	52	36	900
469	5	4705	4169	gi|755152	highly hydrophobic integral membrane protein [Bacillus subtilis] sp|P42953|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG.	52	32	537
495	1	1262	631	gi|1204607	transcription activator [Haemophilus influenzae]	52	25	630
505	7	6004	5762	gi|142440	ATP-dependent nuclease [Bacillus subtilis]	52	28	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
517	2	1162	1614	gi|166162	Bacteriophage phi-11 int gene activator (Staphylococcus acteriophage phi 11)	52	35	453
543	2	444	1295	gi|1215693	putative orf; GT9_orf434 (Mycoplasma pneumoniae)	52	25	852
586	1	1	336	gi|581648	epiB gene product (Staphylococcus epidermidis)	52	36	336
773	1	848	426	gi|1279769	FdhC (Methanobacterium thermoformicicum)	52	30	423
1120	2	100	330	gi|142439	ATP-dependent nuclease (Bacillus subtilis)	52	35	231
1614	1	691	347	gi|289262	comE ORF1 (Bacillus subtilis)	52	28	345
2495	1	1	324	gi|216151	DNA polymerase (gene L; ttg start codon) (Bacteriophage SPO2) gi|579197 SPO2 DNA polymerase (aa 1-648) (Bacteriophage SPO2) pir|A21498|DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2	52	34	324
2931	1	566	285	gi|1256136	YbbG (Bacillus subtilis)	52	30	282
2943	1	577	320	gi|41713	hlsA ORP (AA 1-245) (Escherichia coli)	52	35	258
2993	1	588	295	gi|298032	EF (Streptococcus suis)	52	34	294
3667	1	612	307	gi|849025	hypothetical 64.7-kDa protein (Bacillus subtilis)	52	36	306
3944	1	478	260	gi|1218040	BAA (Bacillus licheniformis)	52	36	219
3954	2	613	347	gi|894064	U87 (Human herpesvirus 6)	52	36	267
3986	1	90	401	gi|1205919	Na+ and Cl- dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	52	90	112
4002	1	3	389	gi|40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p|P23129|ODO1_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA- KETOGLUTARATE DEHYDROGENASE)	52	42	367
4020	1	1	249	gi|159388	ornithine decarboxylase [Leishmania donovani]	52	47	249
4098	1	438	220	gi|409795	No definition line found [Escherichia coli]	52	32	219
4248	1	3	212	gi|965077	Adr4p [Saccharomyces cerevisiae]	52	40	210
7	1	3	575	gi|889747	putative cel operon regulator [Bacillus subtilis]	51	28	573
21	4	2479	3276	gi|1510962	indole-3-glycerol phosphate synthase [Methanococcus jannaschii]	51	32	798
22	9	5301	5966	gi|1303933	YqlN [Bacillus subtilis]	51	25	666
43	3	1516	1283	gi|1519460	srp1 [Schizosaccharomyces pombe]	51	31	234
44	17	11042	11305	gi|42011	moaD gene product [Escherichia coli]	51	35	264
51	11	6453	6731	gi|495471	vacuolating toxin [Helicobacter pylori]	51	37	279

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
52	4	2517	2995	gi|1256652	25% identity to the E.coli regulatory protein MprA; putative (Bacillus subtilis)	51	32	459
57	10	7331	6843	gi|508173	EIIA domain of PTS-dependent Gat transport and phosphorylation Escherichia coli)	51	32	489
59	1	29	1111	gi|299163	alanine dehydrogenase (Bacillus subtilis)	51	33	1083
67	20	15791	16576	gi|1510977	M. jannaschii predicted coding region MJ0938 (Methanococcus jannaschii)	51	24	786
69	2	1559	1218	gi|467359	unknown (Bacillus subtilis)	51	34	342
71	1	3	1196	gi|298032	EF (Streptococcus suis)	51	32	1194
78	2	349	176	gi|1161242	proliferating cell nuclear antigen (Styela clava)	51	28	174
99	4	3357	4040	gi|642795	TFIID subunit TAFII55 (Homo sapiens)	51	25	684
109	1	2852	1428	gi|580920	rodD (gtaA) polypeptide (AA 1-673) (Bacillus subtilis) pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13484|TAGB_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROTEIN B).	51	27	1425
109	9	6007	6693	gi|1204815	hypothetical protein (SP:P32662) (Haemophilus influenzae)	51	23	687
112	1	1066	2352	pir|S05330|S05	maltose-binding protein precursor - Enterobacter aerogenes	51	42	1287
112	13	14432	12855	gi|405857	yehU (Escherichia coli)	51	29	1578
114	9	9725	8967	gi|435098	orf1 (Mycoplasma capricolum)	51	30	759
115	1	1	912	gi|141110	ORF YIhL085w (Saccharomyces cerevisiae)	51	23	912
127	10	9647	10477	gi|1204314	H. influenzae predicted coding region HI0056 (Haemophilus influenzae)	51	37	831
152	9	6814	7356	gi|431929	hunI regulatory protein (Mycoplasma sp.)	51	38	543
154	2	575	1153	gi|1237044	unknown (Mycobacterium tuberculosis)	51	36	579
154	7	6587	5634	gi|1409286	bmrU (Bacillus subtilis)	51	27	954
171	8	6943	6236	gi|1205484	hypothetical protein (SP:P13918) (Haemophilus influenzae)	51	32	708
184	1	1	291	gi|466886	B1495_C3_206 (Mycobacterium leprae)	51	33	291
212	5	1501	2139	pir|A45605|A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	51	23	639
228	2	707	1378	gi|8204	nuclear protein (Drosophila melanogaster)	51	27	672
236	8	8137	7481	gi|49272	Asparaginase (Bacillus licheniformis)	51	31	657
243	4	4637	3546	gi|1511102	mevalonate kinase (Methanococcus jannaschii)	51	29	1092

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
257	4	3540	3373	gi|:204579	H. influenzae predicted coding region H10326 [Haemophilus influenzae]	51	22	168
258	3	2397	1609	gi|160299	glutamic acid-rich protein [Plasmodium falciparum] pir|A54514|A54514 glutamic acid-rich protein precursor - Plasmodium falciparum	51	34	789
265	5	2419	3591	gi|580841	[F1 (Bacillus subtilis)]	51	32	1172
298	2	518	748	gi|1336162	SCPB [Streptococcus agalactiae]	51	34	231
316	9	5817	7049	gi|413953	1pa-29d gene product [Bacillus subtilis]	51	39	1233
332	2	3775	2057	gi|1209012	mutS [Thermus aquaticus thermophilus]	51	26	1719
364	4	3816	4991	gi|528991	unknown [Bacillus subtilis]	51	32	1176
440	2	448	684	gi|2819	transferase (GAL10) (AA 1 - 687) [Kluyveromyces lactis] r|S01407|XUVKG UDPglucose 4-epimerase (EC 5.1.3.2) - yeast uyveromyces marxianus var. lactis)	51	32	237
495	2	1353	1177	gi|297861	protease G [Erwinia chrysanthemi]	51	41	177
495	3	2287	1718	gi|1513317	serine rich protein [Entamoeba histolytica]	51	25	570
506	1	840	421	gi|455320	cII protein [Bacteriophage P4]	51	33	420
600	1	1474	983	gi|587532	orf, len: 201. CAI: 0.16 [Saccharomyces cerevisiae] pir|S48818|S48818 hypothetical protein - yeast (Saccharomyces erevisiae)	51	30	492
607	3	479	934	gi|1511524	hypothetical protein (SP:P17002) [Methanococcus jannaschii]	51	40	456
686	2	127	600	gi|493017	endocarditis specific antigen [Enterococcus faecalis]	51	30	474
726	1	33	230	gi|1353851	unknown [Prochlorococcus marinus]	51	45	198
861	1	176	652	gi|410145	dehydroquinate dehydratase [Bacillus subtilis]	51	34	477
869	1	782	393	gi|40100	rodC (tagI) polypeptide (AA 1-746) [Bacillus subtilis] ir|S06049|S06049 rodC protein - Bacillus subtilis p|P13485|TAGF_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F.	51	23	390
1003	1	642	322	gi|1279707	hypothetical phosphoglycerate mutase [Saccharomyces cerevisiae]	51	39	321
1046	2	866	624	gi|510257	glycosyltransferase [Escherichia coli]	51	29	243
1467	1	702	352	gi|1511175	M. jannaschii predicted coding region MJ1177 [Methanococcus jannaschii]	51	32	351
2558	1	457	230	sp|P10582|DPOM_	DNA POLYMERASE (EC 2.7.7.7) (S-1 DNA ORF 3).	51	26	228
3001	1	779	399	gi|809543	cbrC protein [Erwinia chrysanthemi]	51	27	381
3604	1	1	399	pir|JC4210|JC42	)3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - mouse	51	37	399
3732	1	2	316	gi|145906	acyl-CoA synthetase [Escherichia coli]	51	33	315

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3791	1	2	274	gi|1061351	semaphorin III family homolog [Homo sapiens]	51	37	273
3995	1	46	316	gi|216346	surfactin synthetase [Bacillus subtilis]	51	38	291
4193	1	612	307	gi|42749	ribosomal protein L12 (AA 1-179) [Escherichia coli] [r|S04776|XECPL peptide N-acetyltransferase rimL (EC 2.3.1.-) - cherichia coli]	51	25	306
4539	1	367	185	gi|1408494	homologous to penicillin acylase [Bacillus subtilis]	51	40	183
4562	1	442	239	gi|1458280	coded for by C. elegans cDNA cm01e7; Similar to hydroxymethylglutaryl-CoA synthase [Caenorhabditis elegans]	51	35	204
1	4	3576	4859	gi|559160	GRAIL score: null; cap site and late promoter motifs present patresm; putative [Autographa californica nuclear polyhedrosis irus]	50	44	1284
11	7	4044	5165	gi|1146207	putative [Bacillus subtilis]	50	35	1122
11	13	10509	9496	gi|1208451	hypothetical protein [Synechocystis sp.]	50	39	1014
19	1	2034	1018	gi|413966	ipa-42d gene product [Bacillus subtilis]	50	29	1017
20	11	8586	8407	gi|1323159	ORF YGR103w [Saccharomyces cerevisiae]	50	28	180
24	5	5408	4824	gi|496280	structural protein [Bacteriophage Tuc2009]	50	29	585
34	4	1926	2759	gi|1301966	YqjO [Bacillus subtilis]	50	36	834
38	30	22865	23440	gi|1072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]	50	32	576
47	2	1705	2976	gi|153015	FemA protein (Staphylococcus aureus)	50	29	1272
56	13	15290	15841	gi|606096	ORF [167; end overlaps end of o100 by 14 bases; start overlaps f174, ther starts possible [Escherichia coli]	50	30	552
57	1	2135	1077	gi|640922	xylitol dehydrogenase [unidentified hemiascomycete]	50	29	1059
58	2	628	1761	gi|143725	putative [Bacillus subtilis]	50	29	1134
88	6	4393	3884	gi|1072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) [Caenorhabditis elegans]	50	32	510
89	5	3700	3356	gi|1276658	ORF174 gene product (Porphyra purpurea)	50	25	345
141	1	3	239	gi|476024	carbamoyl phosphate synthetase II [Plasmodium falciparum]	50	33	237
151	1	186	626	gi|1403441	unknown [Mycobacterium tuberculosis]	50	35	441
166	7	11065	9623	gi|895747	putative ccl operon regulator [Bacillus subtilis]	50	32	1443
201	6	5284	5096	gi|160229	circumsporozoite protein [Plasmodium reichenowi]	50	42	189
206	22	30784	29555	gi|1052754	LmrP integral membrane protein [Lactococcus lactis]	50	24	1230

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
211	4	1523	1927	gi|410131	ORFX7 [Bacillus subtilis]	50	29	405
214	4	2411	3295	sp|P37348|YECE_	HYPOTHETICAL PROTEIN IN ASPS 5'REGION (FRAGMENT).	50	37	885
228	7	5068	4406	gi|313580	envelope protein [Human immunodeficiency virus type 1] pir|S35835|S35835 envelope protein - human immunodeficiency virus ype 1 (fragment) (SUB 1-77)	50	35	663
272	2	3048	1723	gi|1408485	B65G gene product [Bacillus subtilis]	50	22	1326
273	2	1616	984	gi|384186	[phosphoglycerate mutase (Saccharomyces cerevisiae]	50	28	633
328	2	2507	1605	gi|148896	lipoprotein [Haemophilus influenzae]	50	26	903
332	4	5469	3802	gi|1526547	DNA polymerase family X [Thermus aquaticus]	50	27	1668
342	5	3473	1931	gi|456562	G-box binding factor [Dictyostelium discoideum]	50	35	459
352	1	1478	741	gi|288301	ORF2 gene product [Bacillus megaterium]	50	29	738
40H	7	5299	5523	gi|11665	ORF2116 [Marchantia polymorpha]	50	27	225
420	3	650	1825	gi|757842	UDP-sugar hydrolase [Escherichia coli]	50	30	1176
464	1	1	591	gi|487282	Na+ -ATPase subunit J [Enterococcus hirae]	50	29	591
472	2	1418	864	gi|551875	BglR [Lactococcus lactis]	50	23	555
520	1	23	541	gi|567016	CapE [Staphylococcus aureus]	50	27	519
529	1	6	410	gi|1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	50	34	405
534	5	7726	6059	gi|295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA ependant RNA polymerase I and III [Saccharomyces cerevisiae]	50	18	1668
647	1	2990	1497	gi|405568	TraI protein shares sequence similarity with a family of opoisomerases [Plasmid pSK41]	50	31	1494
664	3	1133	711	gi|410007	leukocidin P component [Staphylococcus aureus, MRSA No. 4, Peptide, 23 aa]	50	32	423
678	1	1	627	gi|238032	EP [Streptococcus suis]	50	29	627
755	3	947	1171	gi|150572	cytochrome c1 precursor [EC 1.10.2.2] (Paracoccus denitrificans] gi|45465 cytochrome c1 (AA 1-450) [Paracoccus denitrificans] pir|C29413|C29413 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) ytochrome c1 precursor - Paracoccus denitrificans sp|P15627|CY1	50	37	225
827	1	1163	683	gi|142020	heterocyst differentiation protein [Anabaena sp.]	50	21	681
892	1	3	752	gi|408485	B65G gene product [Bacillus subtilis]	50	27	750
910	2	418	887	gi|L04727	tyrosine-specific transport protein [Haemophilus influenzae]	50	25	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
933	1	524	760	gi|1205451	cell division inhibitor [Haemophilus influenzae]	50	32	237
973	1	424	236	gi|886947	orf3 gene product [Saccharomyces cerevisiae]	50	40	189
1009	1	653	429	gi|153727	M protein [group G streptococcus]	50	28	225
1027	1	511	257	gi|413934	ipa-10r gene product [Bacillus subtilis]	50	25	255
1153	2	556	326	gi|773676	nccA [Alcaligenes xylosoxydans]	50	36	231
1222	1	798	400	gi|1408485	B65G gene product [Bacillus subtilis]	50	21	399
1350	1	692	399	gi|289272	ferrichrome-binding protein [Bacillus subtilis]	50	32	294
2945	1	366	184	gi|171704	hexaprenyl pyrophosphate synthetase (COQ1) [Saccharomyces erevisiae]	50	34	183
2968	2	1604	804	gi|397526	clumping factor [Staphylococcus aureus]	50	33	801
2998	2	657	394	gi|495696	F54E7.3 gene product [Caenorhabditis elegans]	50	40	264
3046	2	506	306	pir|S13819|S138	acyl carrier protein - Anabaena variabilis (fragment)	50	32	201
3063	1	547	275	gi|174190	iucA gene product [Escherichia coli]	50	29	273
3174	1	3	146	gi|151900	alcohol dehydrogenase [Rhodobacter sphaeroides]	50	31	144
3792	1	625	314	gi|1001423	hypothetical protein [Synechocystis sp.]	50	35	312
3800	1	2	262	gi|114733	NAD-dependent beta-hydroxybutyryl coenzyme A dehydrogenase Clostridium acetobutylicum	50	28	261
3946	1	373	188	gi|576765	cytochrome b [Myrmecia pilosula]	50	38	186
3984	1	578	291	sp|P37348|YECE_	HYPOTHETICAL PROTEIN IN ASPS 5'REGION (FRAGMENT)	50	37	288
37	10	8250	7885	gi|1204367	hypothetical protein (GB:U14003_278) [Haemophilus influenzae]	49	30	366
46	16	11802	14848	r|466860	acd: B1308_F1_34 [Mycobacterium leprae]	49	24	1047
59	5	2267	3601	p|606304	ORF_o462 [Escherichia coli]	49	27	1335
112	18	17884	18615	gi|559502	ND4 protein (AA 1 - 409) [Caenorhabditis elegans]	49	25	732
138	9	6973	7902	gi|303953	esterase [Acinetobacter calcoaceticus]	49	29	930
217	6	4401	5138	gi|496254	fibronectin/fibrinogen-binding protein [Streptococcus pyogenes]	49	31	738
220	12	11803	12657	gi|397526	clumping factor [Staphylococcus aureus]	49	31	855
228	4	1842	2492	pir|S23692|S236	hypothetical protein 9 - Plasmodium falciparum	49	24	651
268	1	5016	2614	gi|143047	ORFB [Bacillus subtilis]	49	26	2403

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
271	2	1164	1373	gi|1001257	hypothetical protein (Synechocystis sp.)	49	38	210
300	3	4340	3180	gi|1510796	hypothetical protein (GP:X91006_2) (Methanococcus jannaschii)	49	26	1161
381	1	2281	1142	gi|396301	matches PS00041: Bacterial regulatory proteins, araC family signature (Escherichia coli)	49	29	1140
466	1	3	947	gi|1303863	YqgP (Bacillus subtilis)	49	26	945
666	1	379	191	gi|633112	ORF1 (Streptococcus sobrinus)	49	29	189
670	2	403	1014	gi|1122758	unknown (Bacillus subtilis)	49	32	612
709	1	1433	795	gi|143830	xpaC (Bacillus subtilis)	49	29	639
831	1	943	473	gi|401786	phosphomannomutase (Mycoplasma pirum)	49	29	471
1052	1	422	213	gi|1303799	YqeN (Bacillus subtilis)	49	21	210
1800	1	142	172	gi|216300	peptidoglycan synthesis enzyme (Bacillus subtilis) sp|P37585|MURG_BACSU MURG PROTEIN UPD-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-PENTAPEPTIDE)PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE RANSFERASE).	49	28	171
2430	1	2	376	sp|P27434|YFGA_	HYPOTHETICAL 36.2 KD PROTEIN IN NDK-OCPE INTERGENIC REGION.	49	26	375
3096	1	542	273	gi|516360	surfactin synthetase (Bacillus subtilis)	49	25	270
32	4	3771	3100	gi|1217963	hepatocyte nuclear factor 4 gamma (HNF4gamma) (Homo sapiens)	48	36	672
38	1	1	609	gi|1205790	H. influenzae predicted coding region HI1555 (Haemophilus influenzae)	48	28	609
45	6	5021	6427	gi|1524267	unknown (Mycobacterium tuberculosis)	48	20	1407
59	14	16346	1096	gi|1197336	Lmp3 protein (Mycoplasma hominis)	48	28	14751
61	1	3	608	gi|1511555	quinolone resistance norA protein protein (Methanococcus jannaschii)	48	30	606
61	3	3311	3646	gi|1303893	YqhL (Bacillus subtilis)	48	29	336
114	1	98	415	gi|1671708	su(s) homolog; similar to Drosophila melanogaster suppressor of able (su(s)) protein. Swiss-Prot Accession Number P22293 Drosophila virilis)	48	25	318
121	1	1131	610	gi|1314584	unknown (Sphingomonas S88)	48	29	522
136	1	2014	1280	gi|1205968	H. influenzae predicted coding region HI1738 (Haemophilus influenzae)	48	23	735
171	10	8220	9557	gi|1208454	hypothetical protein (Synechocystis sp.)	48	34	1338
175	1	3625	1814	gi|396400	similar to eukaryotic Na+/H+ exchangers (Escherichia coli) sp|P12703|YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS INTERGENIC REGION (O549).	48	29	1812
194	1	2	385	gi|1510493	M. jannaschii predicted coding region MJ0419 (Methanococcus jannaschii)	48	25	384

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
197	1	901	452	gi|1045714	spermidine/putrescine transport ATP-binding protein [Mycoplasma genitalium]	48	25	450
203	1	1	396	gi|940288	protein localized in the nucleoli of pea nuclei; ORF; putative Pisum sativum]	48	29	396
204	1	1163	698	gi|529202	No definition line found [Caenorhabditis elegans]	48	25	666
206	20	34815	27760	gi|511490	gramicidin S synthetase 2 [Bacillus brevis]	48	27	7056
212	1	2	166	gi|295899	nucleolin [Xenopus laevis]	48	14	165
220	10	12652	11426	gi|44073	SecY protein [Lactococcus lactis]	48	23	1227
243	6	6450	5491	gi|1184118	mevalonate kinase [Methanobacterium thermoautotrophicum]	44	30	960
264	4	5434	3308	gi|1015903	ORF YJR151c [Saccharomyces cerevisiae]	48	26	2127
441	1	1532	768	gi|142863	replication initiation protein [Bacillus subtilis] pir|B26580|B26580 replication initiation protein - Bacillus subtilis	48	23	765
444	5	3898	5298	pir|145816	putative [Escherichia coli]	48	24	1401
484	2	388	1110	gi|46551	transmembrane protein (kdpD) [Escherichia coli]	48	18	723
542	3	1425	2000	pir|S28969|S289	N-carbamoylsarcosine amidohydrolase (EC 3.5.1.59) - Arthrobacter sp.	48	27	576
566	1	3	1019	gi|53490	tetracenomycin C resistance and export protein [Streptomyces laucescens]	48	24	1017
611	1	2	730	gi|1103507	unknown [Schizosaccharomyces pombe]	48	38	729
624	1	1255	665	gi|144859	ORF B [Clostridium perfringens]	48	26	591
846	1	1014	508	gi|537506	paramyosin [Dirofilaria immitis]	48	27	507
1020	1	66	950	gi|1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	48	30	885
1227	1	1	174	gi|493720	lipoxygenase [Pisum sativum]	48	35	174
1266	1	1	405	gi|882452	ORF_[211; alternate name yggA; orf5 of X14436 [Escherichia coli] gi|41425 ORF5 (AA 1-197) [Escherichia coli] (SUB 15-211)	48	24	405
2071	1	707	381	gi|1408486	HS74A gene product [Bacillus subtilis]	48	25	327
2398	1	463	233	gi|1500401	reverse gyrase [Methanococcus jannaschii]	48	40	231
2425	1	476	246	pir|H48563|H485	G1 protein - fowlpox virus (strain HP444) (fragment)	48	40	231
2432	1	446	225	gi|1353703	Trio [Homo sapiens]	48	33	222
2453	1	794	399	gi|142850	division initiation protein [Bacillus subtilis]	48	29	396
2998	1	469	236	gi|577569	PepV [Lactobacillus delbrueckii]	48	31	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
3042	1	14	280	gi|945219	mucin [Homo sapiens]	48	35	267
3686	1	1	405	gi|145836	putative [Escherichia coli]	48	25	405
4027	2	492	301	pir.S51177|S11	trans-activator protein - Equine infectious anemia virus	48	32	192
4	2	3641	2232	gi|1303989	Yqk1 [Bacillus subtilis]	47	24	1410
24	2	599	1084	gi|540081	PC4-1 gene product [Bradysia hygida]	47	28	486
36	10	7524	6925	gi|1209923	esterase [Acinetobacter lwoffii]	47	26	600
43	2	196	1884	gi|1403455	unknown [Mycobacterium tuberculosis]	47	27	1689
44	22	16118	15108	gi|1511555	quinolone resistance norA protein protein [Methanococcus jannaschii]	47	31	1011
69	?	7141	6710	gi|1438466	Possible operon with orfG. Hydrophilic, no homologue in the atabase; putative [Bacillus subtilis]	47	29	432
81	4	5022	4279	gi|466882	ppsl; B1496_C2_189 [Mycobacterium leprae]	47	24	744
120	12	9135	8863	gi|927340	D9509.27p; CAI; 0.12 [Saccharomyces cerevisiae]	47	38	273
142	1	2022	1174	gi|486143	ORF YKL094w [Saccharomyces cerevisiae]	47	32	849
168	1	2178	1093	gi|1177254	hypothetical EceB protein [Bacillus subtilis]	47	29	1086
263	1	1884	943	gi|142822	D-alanine racemase cds [Bacillus subtilis]	47	34	942
279	1	1109	561	gi|516608	2 predicted membrane helices, homology with B. subtilis men Orf3 Rowland et. al. unpublished Accession number M74103), approximately 1 minutes on updated Rudd map; putative [Escherichia coli] sp|P37355|YFBB_ECOLI HYPOTHETICAL 26.7 KD PROTEIN IN MEND-MENB	47	31	549
345	2	2620	1676	gi|1204835	hippuricase [Haemophilus influenzae]	47	28	945
389	2	152	400	gi|456562	G-box binding factor [Dictyostelium discoideum]	47	32	249
391	1	1	831	gi|1420856	myo-inositol transporter [Schizosaccharomyces pombe]	47	19	831
404	3	2072	2773	gi|1255425	C33G8.2 gene product [Caenorhabditis elegans]	47	17	702
529	5	2145	3107	gi|1303973	Yqjv [Bacillus subtilis]	47	29	963
565	2	2321	1257	gi|142824	processing protease [Bacillus subtilis]	47	28	1065
654	1	962	483	gi|243353	ORF 5' of ECRF3 [herpesvirus saimiri HVS, host-squirrel monkey, eptide, 407 aa]	47	23	480
692	1	115	633	gi|150756	40 kDa protein [Plasmid pJM1]	47	25	519
765	1	1614	819	gi|1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	47	28	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	start (nt)	stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
825	2	211	1023	gi|397526	clumping factor (Staphylococcus aureus)	47	32	813
914	1	1	615	gi|558073	polymorphic antigen (Plasmodium falciparum)	47	29	615
1076	1	1	753	gi|1147557	Aspartate aminotransferase (Bacillus circulans)	47	33	753
1351	1	793	398	gi|755153	ATP-binding protein (Bacillus subtilis)	47	20	396
4192	1	3	293	gi|45836	putative (Escherichia coli)	47	24'	291
5	6	4708	4361	gi|305080	myosin heavy chain (Entamoeba histolytica)	46	30	348
11	4	2777	3058	gi|603639	Yel040p (Saccharomyces cerevisiae)	46	28	282
46	11	10518	10300	gi|1246901	ATP-dependent DNA ligase (Candida albicans)	46	28	219
61	4	3941	7930	gi|298032	EF (Streptococcus suis)	46	35	3990
132	4	5028	4093	gi|1511057	hypothetical protein SP:P45869 (Methanococcus jannaschii)	46	25	936
170	4	4719	3652	pir|S53910|S519	C4 protein - Sauroleishmania tarentolae	46	26	1068
191	7	9543	8284	gi|1041334	F54D5.7 (Caenorhabditis elegans)	46	25	1260
253	1	1	396	gi|1204449	dihydrolipoamide acetyltransferase (Haemophilus influenzae)	46	35	396
264	3	437	973	gi|180189	cerebellar-degeneration-related antigen (CDR34) [Homo sapiens] gi|182737 cerebellar degeneration-associated protein [Homo sapiens] pir|A29770|A29770 cerebellar degeneration-related protein - human	46	29	537
273	1	485	285	gi|607573	envelope glycoprotein C2V3 region [Human immunodeficiency virus type]	46	35	201
350	1	3	563	gi|537052	ORF_f286 (Escherichia coli)	46	35	561
383	1	2	862	gi|1221884	(urea?) amidolyase (Haemophilus influenzae)	46	31	861
410	4	1876	2490	gi|1110518	proton antiporter efflux pump (Mycobacterium smegmatis)	46	24	615
432	1	2663	1455	gi|1197634	orf4: putative transporter: Method: conceptual translation supplied by author (Mycobacterium smegmatis)	46	27	1209
458	1	2419	1211	gi|15470	portal protain (Bacteriophage SPP1)	46	30	1209
517	5	2477	4192	gi|1523812	orf5 (Bacteriophage A2)	46	23	1716
540	3	1512	1285	gi|215635	pacA (Bacteriophage P1)	46	30	228
587	2	649	1242	gi|537148	ORF_f181 (Escherichia coli)	46	29	594
1218	1	747	391	gi|1205456	single-stranded-DNA-specific exonuclease (Haemophilus influenzae)	46	30	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3685	1	1	402	gi|450688	hsdM gene of EcoprrI gene product (Escherichia coli) pir|S38437|S38437 hsdM protein - Escherichia coli pir|S09629|S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	46	33	402
4176	1	673	338	gi|351460	PIK-C.1 gene product (Xenopus laevis)	46	31	336
37	7	4813	5922	gi|606064	ORF_f408 (Escherichia coli)	45	24	1110
38	16	11699	12004	gi|452192	protein tyrosine phosphatase (PTP-BAS, type 2) (Homo sapiens)	45	24	306
87	2	1748	2407	gi|1064813	homologous to sp:PHOR_BACSU (Bacillus subtilis)	45	23	660
103	12	14182	13385	gi|1001307	hypothetical protein (Synechocystis sp.)	45	22	798
112	14	14791	13811	gi|1204389	H. influenzae predicted coding region HI0131 (Haemophilus influenzae)	45	23	981
145	4	4483	3462	gi|220578	open reading (frame (Mus musculus)	45	20	1023
170	6	6329	4965	gi|238657	AppCcytochrome d oxidase, subunit I homolog (Escherichia coli, K12, peptide, 514 aa)	45	27	1365
206	2	5230	4346	gi|1222056	aminotransferase (Haemophilus influenzae)	45	27	885
228	1	60	716	gi|160299	glutamic acid-rich protein (Plasmodium falciparum) pir|A54314|A54314 glutamic acid-rich protein precursor - Plasmodium alciparum	45	23	657
288	1	2	1015	gi|255425	C13G8.2 gene product (Caenorhabditis elegans)	45	23	1014
313	3	4339	3128	gi|581140	NADH dehydrogenase (Escherichia coli)	45	30	1212
332	1	914	459	gi|870966	P47A4.2 (Caenorhabditis elegans)	45	20	456
344	1	3	221	gi|171225	kinesin-related protein (Saccharomyces cerevisiae)	45	26	219
441	2	1501	1073	gi|142863	replication initiation protein (Bacillus subtilis) pir|B26580|B26580 replication initiation protein - Bacillus subtilis	45	27	429
672	1	2	982	gi|1511334	M. jannaschii predicted coding region MJ1323 (Methanococcus jannaschii)	45	22	981
763	3	1345	851	gi|606180	ORF_f310 (Escherichia coli)	45	24	495
886	3	379	846	gi|726426	similar to protein kinases and C. elegans proteins F37C12.8 and 37C12.5 (Caenorhabditis elegans)	45	30	468
948	1	3	473	gi|156400	myosin heavy chain (isozyme unc-54) (Caenorhabditis elegans) pir|A33958|MMKM myosin heavy chain B - Caenorhabditis elegans sp|P02566|MYSB_CAEEL MYOSIN HEAVY CHAIN B (MHC B)	45	25	471
1158	1	2	376	gi|441155	transmission-blocking target antigen (Plasmodium falciparum)	45	35	375
2551	1	4	285	gi|1276705	ORF287 gene product (Porphyra purpurea)	45	28	282
3967	1	42	374	gi|976025	HrsA (Escherichia coli)	45	28	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
52	7	6931	5846	gi|467378	unknown [Bacillus subtilis]	44	22	1086
138	8	6475	6849	gi|173028	thioredoxin II [Saccharomyces cerevisiae]	44	28	375
221	5	7032	5617	gi|53490	tetracenomycin C resistance and export protein [Streptomyces lauceescens]	44	21	1416
252	2	1331	1122	gi|204989	hypothetical protein (GB:U00022_9) [Haemophilus influenzae]	44	30	210
263	2	3265	2093	gi|136221	carboxypeptidase [Sulfolobus solfataricus]	44	26	1173
365	4	4963	3524	gi|296822	orfl gene product [Lactobacillus helveticus]	44	31	1440
543	3	1315	1833	gi|U63250	low homology to P20 protein of Bacillus lichiniformis and bleomycin acetyltransferase of Streptomyces verticillus [Bacillus subtilis]	44	24	519
544	4	3942	4892	gi|951460	FIM-C.1 gene product [Xenopus laevis]	44	32	951
792	1	1224	613	gi|205680	high molecular weight neurofilament [Rattus norvegicus]	44	28	612
44	18	11303	11911	gi|1511614	molybdopterin-guanine dinucleotide biosynthesis protein A [Methanococcus jannaschii]	43	27	609
59	8	3665	5128	gi|53490	tetracenomycin C resistance and export protein [Streptomyces lauceescens]	43	21	1464
59	10	5516	7527	gi|53022	lipase [Staphylococcus epidermidis]	43	22	1992
99	1	1346	681	gi|1419051	unknown [Mycobacterium tuberculosis]	43	21	666
110	8	9402	12134	gi|397526	clumping factor [Staphylococcus aureus]	43	21	2733
412	3	2782	2303	pir|A60540|A605	sporozoite surface protein 2 - Plasmodium yoelii (fragment)	43	29	480
519	3	2547	3122	sp|Q06530|DHSU_	SULFIDE DEHYDROGENASE (FLAVOCYTOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR [EC 1.8.2.-] (FC) (FCSD).	43	23	576
4	13	12053	13321	gi|295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA ependent RNA polymerase I and III [Saccharomyces cerevisiae]	42	18	1269
94	2	1768	1091	gi|501027	ORF2 [Trypanosoma bruci]	42	31	678
127	4	5791	4550	gi|42029	ORF1 gene product [Escherichia coli]	42	21	1242
297	3	1515	1016	gi|142790	ORF1; putative [Bacillus firmus]	42	25	480
344	6	4097	3525	c|140320	ORF 2 (AA 1-203) [Bacillus thuringiensis]	42	30	573
512	1	2167	1115	gi|405957	yeeF [Escherichia coli]	42	23	1053
631	1	2434	1223	gi|580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir|S06048|S06048 probable rodD protein - Bacillus subtilis sp|P13484|TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPMA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROTEIN E).	42	24	1212

TABLE 2

TABLE 2

C. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match acession	match gene name	% sim	% ident	length (nt)
685	3	2359	1739	gi|1303784	VqeD [Bacillus subtilis]	42	19	621
4132	1	787	395	gi|1022910	protein tyrosine phosphatase [Dictyostelium discoideum]	42	25	393
86	2	1375	884	gi|309506	spermidine/spermine N1-acetyltransferase [Mus saxicola] pir|S43430|S43430 spermidine/spermine N1-acetyltransferase - spiny ouse [Mus saxicola]	41	30	492
191	12	14797	14075	gi|1124957	orf4 gene product [Methanosarcina barkeri]	41	22	723
212	6	2150	3127	gi|15873	observed 35.2Kd protein [Mycobacteriophage 15]	41	26	978
213	3	1261	2000	gi|633692	TrsA [Yersinia enterocolitica]	41	18	738
408	4	2625	3386	gi|1197634	orf4; putative transporter; Method: conceptual translation supplied by author [Mycobacterium smegmatis]	41	24	762
542	1	3	1103	gi|457146	rhoptry protein [Plasmodium yoelii]	41	21	1101
924	1	2	475	pir|JH0148|JH01	nucleolin - rat	41	30	474
1562	1	1	402	gi|552184	asparagine-rich antigen Pfa35-2 [Plasmodium falciparum] pir|S27826|S27826 asparagine-rich antigen Pfa35-2 - Plasmodium falciparum (fragment)	40	20	402
2395	1	518	261	pir|S42251|S422	hypothetical protein 5 - fowlpox virus	40	18	258
4077	1	3	305	gi|1015055	coded for by C. elegans cDNA yk17g1.5; coded for by C. elegans cDNA yk5c9.5; coded for by C. elegans cDNA yk1e9.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans]	39	21	303
958	1	1003	503	gi|1255425	C33G8.2 gene product [Caenorhabditis elegans]	37	25	501
59	12	8294	10636	gi|535260	STARP antigen [Plasmodium reichenowi]	36	24	2343
63	5	3550	8079	gi|298032	EF [Streptococcus suis]	36	19	4530
544	3	2507	3601	gi|1015903	ORF YJR151c [Saccharomyces cerevisiae]	35	22	1095
63	4	1949	3574	gi|552195	circumsporozoite protein [Plasmodium falciparum] sp|P05691|CSP_PLAFL CIRCUMSPOROZOITE PROTEIN (CS) (FRAGMENT).	32	27	1626

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4	1	1234	692
4	3	1712	2278
4	4	3703	1032
4	14	13073	12585
5	2	2539	1601
5	3	1532	1771
5	7	4741	4550
5	9	7939	6422
5	12	8711	8547
6	4	2359	1982
8	1	349	176
11	8	5144	5983
11	9	5968	6498
11	10	6472	6284
11	16	10954	11271
12	5	5352	4942
12	6	4596	4862
15	3	1895	1650
16	10	11263	10835
18	2	1093	917
20	9	9125	7764
20	10	8571	8230
20	12	9201	8803
20	13	12158	10470
23	1	674	339
23	6	6138	5485
23	8	6376	5942

S. aureus - Putative coding regions of novel proteins not similar to known proteins

TABLE 3

Contig ID	ORF ID	Start (nt)	Stop (nt)
23	9	7651	6881
23	15	12618	12830
24	4	4556	4185
24	6	5642	5241
25	2	1824	2402
31	2	505	849
31	3	1177	1524
31	4	2454	3005
32	2	765	1388
32	9	7952	8575
33	10	8591	8728
32	11	9738	9379
32	12	10797	10087
34	2	1315	1049
36	7	5226	5801
36	11	7575	7261
36	12	7424	7621
37	4	3158	2964
38	2	1585	980
38	11	6425	6868
38	20	16982	16371
38	26	20253	20804
38	27	20722	21264
39	1	1	627
40	1	805	404
43	1	796	428
44	4	2674	2324

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
44	5	2484	3263
44	14	10587	10129
44	20	13724	13536
44	21	13596	13994
45	7	6575	6297
46	8	6365	6520
46	12	10449	10976
46	17	15032	15424
47	1	288	1079
48	9	7620	7778
50	1	1612	962
50	2	1621	1316
51	1	738	370
51	5	2520	2245
53	1	442	287
53	7	6705	6319
54	7	9014	8709
55	1	592	326
55	3	1052	786
56	1	1	261
56	3	1551	1228
56	4	1970	1560
56	17	19092	18712
57	4	1694	3521
57	8	5436	5822
58	9	8885	8553
59	3	1366	1509

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
59	6	3026	2802
59	7	3770	3570
59	9	4946	4563
59	11	7518	8378
59	13	10401	16403
62	2	2696	1521
62	11	5440	5757
63	1	1	336
67	1	900	1781
67	2	1774	2610
67	3	2591	3904
67	8	7110	6955
68	1	78	326
70	6	6761	5199
70	11	8935	8645
77	1	1590	1192
79	2	1509	1228
79	3	1411	1791
83	1	2	603
85	9	8300	8653
85	10	8969	8781
86	3	1426	1232
87	8	9187	9366
88	3	1620	1922
89	1	3	161
89	7	5042	4878
91	1	1098	550

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
91	3	3938	3141
92	2	449	928
92	3	1958	1467
92	9	5638	6024
94	1	661	332
94	3	2445	1813
94	4	2583	2197
96	11	10601	11050
99	6	4672	4523
99	7	5014	4784
100	8	7658	7287
102	7	4697	4368
103	3	2496	2035
104	1	2	694
104	2	699	1277
105	1	1235	693
105	3	3233	2655
105	1	3	221
106	3	1209	1355
107	1	1081	542
109	4	4025	3651
109	13	11625	11996
109	14	11981	12268
109	20	17401	17688
110	1	2	760
114	10	8764	9384
116	1	1	309

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig no	ORF ID	Start (nt)	Stop (nt)
116	3	6273	4462
116	8	11049	9976
116	9	10313	10158
120	5	3703	3320
120	6	4270	3869
120	13	9290	9844
121	2	417	569
126	3	1090	818
127	3	2648	3196
127	5	4084	4395
131	6	6773	6438
132	2	715	1695
134	1	2	667
135	2	512	258
135	3	1124	729
138	1	3	152
138	7	6008	6463
140	1	2060	1032
140	2	2019	1533
140	5	2387	2743
142	2	1160	2388
142	7	8830	7586
143	7	7290	6502
144	1	1227	640
146	1	2	511
146	3	502	1350
146	4	3673	2540

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
146	5	2874	3071
147	1	1	339
149	11	3956	3615
149	12	4036	3785
149	13	4507	4145
149	15	4807	4610
149	16	5495	5049
149	18	5739	5491
149	21	7416	7054
149	23	9216	8521
149	24	9681	9106
149	25	10679	9897
150	2	2303	1587
154	3	1795	1508
154	8	6586	6398
154	14	12704	12147
154	15	13531	12803
156	1	315	593
157	3	1183	2232
158	2	1472	1064
159	3	452	808
161	2	876	1808
161	6	4653	4279
161	7	4803	4540
161	8	4896	4717
161	11	5817	5638
163	2	1604	840

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
163	5	2796	2344
163	7	2952	2647
163	9	4905	5132
164	3	1338	1147
166	3	5213	4854
168	4	2500	2868
168	5	3595	4158
170	3	2517	2777
171	2	2277	1450
171	11	12576	11125
172	1	3	278
172	2	1940	1149
173	1	1289	708
173	5	7001	6114
174	2	593	1105
175	3	2552	2890
175	5	3820	3335
175	7	4142	4506
182	4	5477	4986
184	5	6043	5702
188	2	1210	1755
188	4	2647	2994
189	6	2614	3039
190	3	1998	2564
191	1	1	153
191	2	950	669
191	10	11786	13019

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
191	11	12902	12363
192	1	91	426
195	3	2306	1932
195	5	2899	2606
198	2	1016	1591
201	1	170	625
203	2	783	1466
206	6	8930	7815
206	12	13947	13636
206	21	28208	27960
212	2	170	817
212	3	796	1167
212	7	3128	3436
212	9	3749	4075
213	1	1	705
214	2	1076	570
214	6	4064	3738
214	9	6600	6995
214	10	7864	7469
217	1	1927	965
218	1	178	657
218	3	1776	2156
220	2	1851	1369
220	3	3251	2262
220	7	8275	7208
220	8	10244	8661
220	9	11796	10216

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
221	4	3095	2613
221	9	11426	10757
226	1	3	659
226	2	2196	1459
226	3	1476	1961
227	1	2	487
227	2	460	975
227	4	1855	2121
227	5	2052	2345
227	6	4760	3768
227	9	5591	6367
228	5	2503	2877
228	6	2846	3526
233	7	3944	3762
236	2	808	579
238	2	1975	1391
239	2	1417	905
241	5	4495	4334
242	2	1677	1363
243	1	127	576
244	1	1291	647
244	2	3035	1962
245	2	1614	1258
246	1	69	215
246	4	738	1733
249	3	3906	3712
250	1	494	249

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
254	1	1	156
256	2	956	1144
257	3	3700	3227
260	4	4906	4580
261	4	2196	2606
261	6	3214	3681
264	2	155	439
264	5	5252	4533
264	6	4739	5107
267	2	1323	931
268	4	5140	4700
272	1	862	446
272	3	1200	1439
272	9	4691	4909
272	10	6469	6035
276	4	1746	1901
278	1	224	553
278	5	3299	3448
278	7	4849	5127
285	2	551	736
288	3	1756	1950
288	5	2055	2276
289	1	2107	1055
290	2	2234	1932
291	2	332	622
291	5	1545	2051
295	3	1606	1349

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
295	4	2728	2141
295	5	2220	2762
297	2	788	465
298	1	2	205
300	2	2380	1928
301	7	2794	2624
304	1	3	194
306	1	109	654
306	5	4036	4257
307	1	674	339
307	8	3645	3995
308	1	1	654
308	2	1120	599
308	4	2643	2332
313	2	2314	1919
314	1	10	702
316	2	982	1341
316	6	2758	3165
317	1	2	1114
317	3	4570	3458
321	6	5645	5217
321	7	6319	6140
321	8	7450	6794
322	2	827	543
326	2	165	1112
326	3	1117	1467
328	1	936	469

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nc)	Stop (nc)
328	5	3452	3276
329	1	3	719
329	2	781	1212
329	3	1471	1833
330	1	576	289
330	3	1447	1623
332	3	2353	2204
332	7	4971	5338
333	2	3295	3128
335	1	864	433
337	2	95	526
340	2	1658	1356
341	1	3	281
341	3	2476	3192
341	5	3618	3944
341	6	3929	4558
344	5	3197	2889
345	1	1532	768
346	2	221	592
350	3	1410	1598
352	2	2178	1765
352	3	7316	4596
352	7	7967	8404
352	8	8906	9247
352	9	10171	9854
359	1	1	546
362	1	3	656

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
364	2	2158	1808
364	8	10974	10714
365	2	1612	1313
365	5	4680	4090
365	7	4980	6239
366	3	520	1719
367	3	906	1085
368	1	748	494
375	1	2	136
380	3	1351	1097
389	1	1	276
390	1	2	877
390	2	1373	1549
391	2	751	560
395	1	391	197
396	1	2132	1068
398	3	1344	1141
399	1	176	669
401	3	566	847
402	2	100	465
404	8	5561	5370
408	2	3507	2269
408	3	2875	2672
408	5	3524	4423
410	3	2111	1890
413	1	880	488
416	1	607	320

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
416	2	578	847
416	3	2195	1390
417	1	3	179
417	2	161	616
420	2	788	513
422	2	357	677
431	2	856	1407
432	2	446	1084
433	1	1	417
433	3	2311	2033
434	1	942	535
434	2	2089	1235
440	1	1	450
442	2	1269	3320
443	3	1873	1520
444	1	1	696
444	7	6761	6366
451	1	940	614
453	2	896	636
453	8	3833	4786
453	9	4718	4512
453	10	4937	4731
455	1	434	219
455	2	472	930
459	1	265	687
462	1	2	247
466	2	1494	907

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
467	1	654	349
468	1	2	250
469	1	1488	925
469	3	2386	3372
469	4	3464	3706
470	1	77	538
470	6	4098	3694
470	7	6330	5686
470	9	7351	8181
470	10	8175	9773
471	1	940	500
471	2	1562	1017
476	1	70	267
477	1	2	760
477	3	1764	2081
477	4	2066	2332
480	5	4016	4261
481	2	956	480
486	3	613	774
487	6	1795	2112
488	1	715	359
492	1	127	675
493	1	2	520
493	2	496	1242
502	3	1149	1571
504	1	690	346
505	5	4566	4150

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
511	2	1741	1232
512	2	583	747
515	1	609	812
517	4	2179	2511
520	4	2097	2360
520	6	1908	1669
527	1	1	498
528	1	637	335
529	2	1679	1104
530	7	5298	5534
536	1	308	156
538	1	1362	736
53A	3	2203	2AR0
538	5	3531	3121
538	6	4348	3731
540	1	996	664
540	2	1495	1031
541	1	89	433
541	2	719	432
542	2	1048	1272
545	2	1012	734
551	1	2145	1129
555	2	892	704
558	3	1357	1154
558	4	1760	1458
558	5	2105	1821
558	6	2166	2020

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
558	7	2636	2322
558	8	3053	2802
558	9	3986	3453
560	1	473	921
565	3	1706	1485
571	1	308	156
571	3	994	1206
577	1	2	199
577	2	163	453
579	1	1	477
579	2	1784	1200
583	1	1988	996
585	1	946	539
587	1	22	573
588	2	1896	1372
588	3	1742	1554
590	1	47	334
592	2	1455	1141
593	1	2	775
593	2	817	1122
595	1	87	890
596	3	1593	1435
602	1	8	169
603	5	1071	1469
606	1	322	768
607	5	1444	1226
610	1	1029	541

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
612	1	3	500
616	1	991	650
617	2	736	491
622	1	36	347
625	4	2046	2549
627	1	67	210
628	1	901	452
631	3	4789	4004
634	1	1448	759
636	1	189	368
636	2	1929	1063
637	2	2323	1994
638	1	227	1081
639	1	518	261
639	2	1377	811
641	1	118	444
642	3	1615	1331
642	4	2260	1847
643	1	3	608
645	4	1534	1758
645	6	2025	2321
645	7	2940	2688
648	1	2	1045
660	1	77	601
660	2	576	872
661	1	1725	961
664	2	89	304

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
667	1	3	413
668	1	1	330
671	2	812	516
673	1	3	338
674	2	865	584
679	1	1	237
679	3	1589	1906
688	1	1236	835
688	2	1352	1077
694	1	3	143
696	2	818	432
706	1	36.	224
709	3	1183	1449
711	1	3	908
715	1	3	167
716	1	2	637
721	1	133	570
722	1	763	383
723	1	1656	829
723	2	1498	1112
727	1	2	472
729	1	268	441
731	1	130	828
735	1	2	214
736	1	3	782
738	1	2	298
742	1	3	230

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
745	3	1148	780
748	2	282	464
749	1	685	344
751	1	901	452
755	1	97	522
755	2	520	918
758	2	663	400
764	2	1053	746
767	1	1	405
768	1	2	373
771	1	1058	534
778	1	1735	902
785	1	1790	1023
787	1	1260	631
791	1	3	224
799	1	15	260
804	1	304	711
805	1	3	680
808	1	219	842
810	1	2221	1112
810	2	1774	1442
812	1	38	979
817	1	714	358
818	2	487	1104
819	2	1529	1032
819	3	1748	1419
820	1	195	1064

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
828	1	506	255
829	1	48	800
830	1	578	291
832	1	594	298
835	1	320	796
840	3	491	709
845	1	912	457
850	2	103	449
853	1	715	359
860	1	2	256
864	2	18	410
864	2	383	715
864	6	1676	1828
870	1	1	588
873	1	906	454
875	1	584	224
877	1	1661	1020
878	1	981	544
879	1	1567	785
881	1	1	243
882	1	389	604
890	1	2	508
905	1	793	398
906	1	852	544
912	1	373	188
913	1	3	290
913	2	1092	547

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
915	1	6	161
915	2	169	402
921	1	126	386
927	1	1578	808
928	1	2	385
929	1	2	400
932	1	2	400
934	1	1	384
936	1	1052	528
937	1	2	616
945	1	220	645
945	2	649	1242
946	1	1702	950
949	1	1	270
951	1	3	362
955	1	3	143
960	1	723	400
963	1	1	162
965	1	690	346
966	1	1079	606
969	1	3	302
971	1	12	170
974	1	319	161
976	1	692	348
977	1	2	211
982	1	1926	982
984	1	589	296

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
987	1	3	467
993	1	1	525
994	1	920	549
1004	1	557	318
1014	1	624	313
1015	1	2	463
1016	1	288	145
1019	1	1205	660
1022	1	839	474
1024	1	595	299
1024	2	276	431
1030	1	673	338
1032	1	355	179
1040	1	794	399
1043	1	3	269
1044	2	115	399
1047	1	1	159
1051	1	704	354
1051	2	1233	733
1063	1	2	400
1069	1	2	148
1069	2	769	533
1075	1	707	399
1077	1	97	405
1081	1	58	438
1086	1	1	384
1087	2	246	431

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1088	1	3	374
1096	1	474	238
1098	1	1015	509
1100	1	1020	511
1100	2	1520	1158
1101	1	703	353
1102	1	385	194
1107	1	2	580
1114	1	3	422
1115	1	2	268
1119	1	22	267
1129	1	40	342
1112	1	360	181
1133	1	609	376
1144	1	446	225
1147	1	558	280
1153	1	1	153
1154	1	3	818
1159	1	1	330
1161	1	341	186
1164	1	427	254
1171	1	19	240
1171	2	108	299
1183	1	2	379
1195	1	355	179
1196	1	1	189
1200	1	33	197

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1203	2	129	464
1222	2	105	401
1232	1	1	187
1240	1	2	175
1247	1	520	311
1271	1	412	221
1286	1	2	595
1295	1	1	165
1306	1	367	185
1314	2	158	631
1316	1	58	570
1359	1	384	193
1370	1	1	402
1371	1	1	345
1374	1	710	357
1378	1	2	400
1392	1	3	413
1411	1	202	432
1433	1	331	167
1450	1	2	256
1453	1	295	149
1471	1	721	398
1477	1	869	639
1502	1	794	399
1518	1	126	449
1534	1	283	143
1546	1	3	401

TABLE 3

S. auraus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1547	1	506	255
1593	1	3	350
1587	1	3	563
1602	2	170	679
1629	1	1	402
1665	1	468	235
1760	1	625	314
1762	1	3	200
1876	2	119	286
1895	1	2	379
1931	1	798	400
1976	2	715	183
2055	2	252	401
2056	1	331	167
2150	1	523	263
2157	1	794	399
2164	1	564	283
2175	1	218	400
2212	1	492	331
2338	1	732	367
2342	1	3	167
2352	1	330	166
2352	2	622	398
2355	1	47	352
2356	1	679	341
2359	1	301	152
2421	1	296	150

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3046	1	367	185
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	357	190
3075	1	440	222
3080	1	1	285
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	298
3118	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	336	169
3142	1	388	203
3144	1	664	386
3151	1	337	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3046	1	167	185
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	357	190
3075	1	440	222
3080	1	1	205
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	298
3118	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	336	169
3142	1	388	203
3144	1	664	386
3151	1	337	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3568	1	751	377
3595	1	757	380
3618	1	2	238
3618	2	130	402
3622	1	86	358
3622	2	664	398
3642	1	876	439
3649	1	781	398
3651	1	625	314
3664	1	467	637
3674	1	55	402
3677	1	619	311
3704	1	1	402
3726	1	535	269
3765	1	510	256
3779	1	554	357
3794	1	266	135
3794	2	667	377
3796	2	638	375
3801	1	474	262
3806	1	453	298
3807	1	42	389
3815	1	798	400
3827	1	3	120
3842	1	781	392
3853	1	671	399
3855	1	1	324

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3857	1	2	235
3861	1	590	297
3865	1	695	399
3897	1	3	173
3897	2	143	400
3898	2	225	401
3921	2	103	342
3927	1	70	375
3930	1	76	234
3946	2	651	382
3951	2	105	377
3965	1	646	344
3972	1	795	400
3981	1	3	311
3998	1	3	356
4001	1	481	296
4003	1	90	335
4018	1	2	259
4018	2	186	401
4021	1	1	345
4043	1	3	344
4054	1	3	344
4066	1	1	150
4070	1	1	324
4072	2	187	390
4073	1	1	285
4077	2	127	372

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig / ID	ORP ID	Start (nt)	Stop (nt)
4083	1	3	359
4090	1	27	368
4101	1	103	297
4105	1	1	306
4107	1	570	286
4119	1	629	339
4121	1	740	372
4123	1	3	230
4127	1	3	341
4128	1	2	391
4130	1	768	415
4146	1	97	381
4157	2	3	206
4186	1	505	254
4224	2	510	256
4239	1	1	348
4242	1	709	356
4252	1	589	296
4253	1	1	174
4256	1	568	323
4258	2	498	334
4267	1	284	144
4271	1	2	304
4287	1	303	163
4289	1	471	319
4302	1	153	305
4304	2	1	186

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4304	2	96	314
4306	1	2	151
4318	1	576	289
4322	1	5	148
4331	1	439	221
4331	2	528	364
4338	1	728	399
4346	1	671	277
4367	2	117	311
4373	1	2	268
4381	1	574	326
4384	1	614	309
4397	1	9	311
4402	1	1	249
4403	1	606	328
4406	1	3	117
4411	1	2	280
4411	2	697	398
4412	1	2	364
4418	1	3	230
4424	1	601	398
4443	1	427	215
4471	1	643	323
4478	1	540	271
4482	1	50	289
4489	1	601	302
4491	1	12	206

TABLE 3

S. aurous - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4495	1	3	179
4496	1	500	252
4500	1	130	306
4511	1	493	248
4518	1	1	246
4526	1	480	241
4527	1	2	163
4532	1	3	239
4543	1	11	175
4567	1	16	200
4573	1	1	231
4578	1	642	322
4619	1	1	180
4620	1	349	176
4662	1	1	246
4669	1	2	157
4680	1	28	183
4690	1	344	174

Table 4

ORF	SEQ ID NO	BLAST HOMOLOG	Antigenic Regions			
			Region 1	Region 2	Region 3	Region 4
168_6	5192	lipoprotein	36-45	84-103	152-161	176-185
238_1	5193	chrA	21-39	48-58	84-95	232-249
51_2	5194	OppB gene product (B. sub·	20-36	70-79	100-112	121-131
278_3	5195	lipoprotein 1	20-29	59-73	85-97	162-171
276_2	5196	lipoprotein	21-33	65-74	177-186	211-220
45_4	5197	ProX	28-37	59-69	85-100	120-129
315_8	5198	hypothetical protein	45-54	88-97	182-192	243-253
154_15	5199	unknown	31-40	48-58	79-88	95-104
228_3	5200	unknown	25-38	40-52	64-74	80-89
228_6	5201	unknown	29-41	89-101	128-143	173-184
50_1	5202	unknown	21-33	52-61	168-182	197-206
112_7	5203	iron-binding periplasmic	21-31	58-67	92-101	111-120
442_1	5204	unknown	30-39	91-100	122-137	182-192
66_2	5205	unknown	50-59	104-116	127-136	167-182
304_2	5206	Q-binding periplasmic	19-28	48-57	75-84	103-116
44_1	5207	hypothetical protein	27-36	86-95	129-138	192-201
161_4	5208	SphX	27-44	149-161	166-175	201-210
46_5	5209	cmpC (permease)	21-33	61-70	83-92	100-109
942_1	5210	traH [Plasmid pSK41]	83-92	109-118	127-142	
5_4	5211	ORF (S. aureus)	12-22	87-96	111-120	151-160
20_4	5212	peptidoglycan hydrolase (S	24-34	129-138	141-150	161-171
328_2	5213	lipoprotein (H. flu)	81-90	123-133	290-299	
520_2	5214	fibronectin binding protein	44-54	63-79	81-90	95-110
771_1	5215	emm1 gene product (S. py	30-39	65-82	96-106	112-121
999_1	5216	predicted trithorax prot. (D	7-16	120-129	157-166	
853_1	5217	ORF2136 (Marchantia polyr	43-52	88-97	102-111	
287_1	5218	psaA homolog	13-22	28-44	72-82	114-124
288_2	5219	cell wall enzyme	14-23	89-98		
596_2	5220	penicillin binding protein 2b	40-49	59-68	76-87	106-115
217_5	5221	fibronectin/fibrinogen bindi	28-37	40-49	62-71	93-111
217_6	5222	fibronectin/fibrinogen bp	10-19	31-40	54-62	73-92
528_3	5223	myosin cross reactive prot€	4-13	29-47	60-73	90-99
171_11	5224	EF	20-31	91-110		
63_4	5225	penicillin binding protein 2b	12-21	59-68	95-104	
353_2	5226		46-55	62-71		
743_1	5227	29 kDa protein in fimA regi	23-32	68-79	94-103	175-184
342_4	5228	Twitching motility	10-19	48-60	83-92	111-121
69_3	5229	arabinogalactan protein	97-106	132-141	158-167	180-189
70_6	5230	nodulin	36-45	48-57	137-160	179-188
129_2	5231	glycerol diester phosphodie	8-17	41-50	55-74	97-106
58_5	5232	PBP (S. aureus)	26-35	70-79	117-126	152-161
188_3	5233	MHC class II analog (S. aure	72-81	94-103	115-124	136-145
236_6	5234	histidine kinase domain (Dic	24-33	52-67	81-94	106-121
310_8	5235	clumping factor (S. aureus)	59-71	77-86	93-102	118-127
601_1	5236	novel antigen/ORF2 (S. aur	45-54	91-104	108-117	186-195
544_3	5237	ORF YJR151c (S. cerevisae)	76-90	101-111	131-140	154-164
662_1	5238	MHC class II analog (S. aure	22-32	71-80	89-98	114-122
87_7	5239	5' nucleotidase precursor ('	29-45	62-71	105-114	125-137
120_1	5240	B65G gene product (B. sub	102-111			

Table 4

ORF	Antigenic Regions (cont)					
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
168_6	244-272	303-315				
238_1	260-269	291-301	308-317			
51_2	140-152	188-208	211-220	256-266	273-283	
278_3	198-209					
276_2	255-268					
45_4	177-199	221-230	234-243	268-279	284-293	304-313
316_8						
154_15	148-157	177-187	202-211			
228_3	101-119	139-154	166-181			
228_6						
50_1						
112_7	136-149	197-211	218-229	253-273		
442_1	199-210	247-257	264-277	287-309		
66_2						
304_2	178-187	250-259				
44_1						
161_4						
46_5	131-141	162-176	206-215	243-252	264-273	285-294
942_1						
5_4	189-205	230-239	246-264	301-318	340-354	378-387
20_4	202-212	217-234	260-275	314-336	366-373	380-391
328_2						
520_2						
771_1	145-154					
999_1						
853_1						
287_1	154-164					
288_2						
596_2	121-130					
217_5	244-253	259-268	288-297	302-311		
217_6	144-158	174-183	188-197	207-216	226-242	
528_3						
171_11						
63_4						
353_2						
743_1	197-207					
342_4						
69_3	195-211					
70_6	206-215	263-272	291-301	331-340	358-371	390-414
129_2	117-127	141-157	168-183	202-211	222-231	261-270
58_5	184-203	260-269	275-299	330-344	372-381	424-433
188_3						
236_6	138-147	163-172	187-198	244-261	268-278	308-317
310_8	131-140	144-153	177-186	190-199	204-213	216-227
601_1	208-218					
544_3	170-179	184-193	224-235	274-287	327-336	352-361
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 11	Region 12	Region 13	Region 14	Region 15	Region 16
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5	306-315					
942_1						
5_4	393-407	416-426	456-465			
20_4	396-405	410-419	461-481			
328_2						
520_2						
771_1						
999_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6	453-471	506-515				
129_2	296-315					
58_5						
188_3						
236_6	358-377	410-423	428-439	442-457	467-476	480-493
310_8	238-251	256-275	281-290	296-310	314-333	338-347
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 17	Region 18	Region 19	Region 20	Region 21	Region 22
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5						
942_1						
5_4						
20_4						
328_2						
520_2						
771_1						
999_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6						
129_2						
58_5						
188_3						
236_6						
310_8	357-366	370-379	429-438	443-452	478-487	551-560
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 23	Region 24	Region 25	Region 26	Region 27	Region 28
168_6						
238_1						
51_2						
278_3						
276_2						
45_4						
316_8						
154_15						
228_3						
228_6						
50_1						
112_7						
442_1						
66_2						
304_2						
44_1						
161_4						
46_5						
942_1						
5_1						
20_4						
328_2						
520_2						
771_1						
999_1						
853_1						
287_1						
288_2						
596_2						
217_5						
217_6						
528_3						
171_11						
63_4						
353_2						
743_1						
342_4						
69_3						
70_6						
129_2						
58_5						
188_3						
236_6						
310_8	622-632	670-685	708-718	823-836	858-867	877-886
601_1						
544_3						
662_1						
87_7						
120_1						

Table 4

ORF	Antigenic Regions		(cont)
	Region 29	Region 30	
168_6			
238_1			
51_2			
278_3			
276_2			
45_4			
316_8			
154_15			
228_3			
228_6			
50_1			
112_7			
442_1			
66_2			
304_2			
44_1			
161_4			
46_5			
942_1			
5_4			
20_4			
328_2			
520_2			
771_1			
999_1			
853_1			
287_1			
288_2			
596_2			
217_5			
217_6			
528_3			
171_11			
63_4			
353_2			
743_1			
342_4			
69_3			
70_6			
129_2			
58_5			
188_3			
236_6			
310_8			
601_1			
544_3			
662_1			
87_7			
120_1			

Table 4

ORF		BLAST HOMOLOG	Antigenic Regions			
			Region 1	Region 2	Region 3	Region 4
46_1	5241	aldehyde dehydrogenase	8-17	36-52	83-96	112-121
63_4	5242	glycerol ester hydrolase (P.	9-26	57-73	93-107	123-133
174_6	5243	ketopantoate hydroxymeth	71-80	203-212	242-254	265-274
206_16	5244	ornithine acetyltransferase	1-10	34-43	54-63	194-210
267_1	5245	NaH-antiporter protein (E. t	120-129	332-347	398-408	
322_1	5246	acriflavin resistance protein	58-75	153-164	203-231	264-284
415_2	5247	transport ATP-binding prot(	108-126	218-227	298-308	315-334
214_3	5248	2-nitropropane dioxygenase	123-136	216-233	283-292	297-306
587_3	5249	clumping factor	5-14	43-54	59-68	76-95
685_1	5250	signal peptidase	59-68	72-81	86-95	99-108
54_3	5251	fibronectin binding protein I	23-32	37-46	50-59	89-98
54_4	5252	fibronectin binding protein I	43-52	66-75	95-104	147-156
54_5	5253	fibronectin binding protein I	49-60	81-90		
54_6	5254	fibronectin binding protein I	55-71	82-97	139-158	175-186
328_1	5255	lipoprotein (H. flu)	11-20	61-70	96-105	

Table 4

ORF	Antigenic Regions			(cont)		
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
46_1	215-242	333-352	376-385	416-432	471-487	
63_4	145-154	191-202	212-223	245-265	274-283	291-300
174_6						
206_16	239-259	275-284				
267_1						
322_1	298-319	350-359				
415_2	344-353	371-380	395-404	456-465	486-495	518-527
214_3	318-337	365-375				
587_3	106-115	142-151	156-166	173-182	186-198	204-213
685_1	113-122	130-145				
54_3	128-138	185-194	217-226	251-260	268-277	295-305
54_4	175-188	191-200	203-212	220-229		
54_5						
54_6	220-230	287-304	317-326	344-353	364-373	378-387
328_1						

Table 4

ORF	Region 11	Region 12	Region 13	Region 14	Region 15	Region 17
46_1						
63_4	306-315	319-328	366-376	395-420	453-462	467-476
174_6						
206_16						
267_1						
322_1						
415_2	539-555					
214_3						
587_3	217-226	278-287	318-327	332-342	351-360	377-386
685_1						
54_3	316-325	329-345	355-372	387-396	416-425	438-448
54_4						
54_5						
54_6	396-407	427-436	514-531	541-550	569-578	612-622
328_1						

Table 4

ORF	Region 18	Region 19	Region 20	Region 21	Region 22	Region 23
46_1						
63_4	485-500	513-525				
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	396-405	426-442	459-470	485-494	505-514	531-562
685_1						
54_3	455-462	472-491	517-536			
54_4						
54_5						
54_6	639-648	673-681	703-715	723-732	749-760	772-788
328_1						

Table 4

| ORF | Antigenic Regions | | | (cont) | | |
	Region 24	Region 25	Region 26	Region 27	Region 28	Region 29
46_1						
63_4						
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	567-578	584-601	607-840	844-854	858-870	877-886
685_1						
54_3						
54_4						
54_5						
54_6	793-802	811-826	834-848	866-876	893-903	907-918
328_1						

Table 4

| ORF | Antigenic Regions | | (cont) |
	Region 30	Region 31	
46_1			
63_4			
174_6			
206_16			
267_1			
322_1			
415_2			
214_3			
587_3	889-911	927-936	
685_1			
54_3			
54_4			
54_5			
54_6	925-944	951-997	
328_1			

SEQUENCE LISTING

5

(1) GENERAL INFORMATION:

  (i) APPLICANT:
      (A) NAME: Human Genome Sciences, Inc.
10
      (B) STREET: 9410 Key West Avenue
      (C) CITY: Rockville
      (D) STATE: Maryland
      (E) COUNTRY: US
15
      (F) POSTAL CODE: 20850

  (ii) TITLE OF INVENTION: Staphylococcus aureus Poly-
              nucleotides and Sequences
20

  (iii) NUMBER OF SEQUENCES: 5255

  (v) COMPUTER READABLE FORM:
25
      (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
      (B) COMPUTER: HP Vectra 486/33
      (C) OPERATING SYSTEM: MSDOS version 6.2
      (D) SOFTWARE: ASCII Text
30

  (vi) CURRENT APPLICATION DATA:
      (A) APPLICATION NUMBER:
      (B) FILING DATE:
35

  (vii) PRIOR APPLICATION DATA:
      (A) APPLICATION NUMBER: US 60/009,861
      (B) FILING DATE: 05-JAN-1996
40

(2) INFORMATION FOR SEQ ID NO:1:

45

50

55

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 5895 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
TCCATTATGA AGTCACAAGT ACTATAAGCT GCGATGTTAC CAATGTTTTT TAAAATCCCA      60

GTAATAAAAT CAAAAAATAA GTTAAATAAT GTATTCATTT TAAGTCCTCC TTAATAAAGa     120

aaataGGTAA TAATGTAATA GCTTCTATTA TGATGCCTAA TTGAATGAAT TGGGCAAATG     180

GCTCTTTGAT GATAAGTGTG ATAATGAAAA GGGTTAAACT AACAATAATC GCATAATATT     240

TTTTTCGTTT AATAAGTCGC ACAGGAATGG GCTTCTTTTT AGTTGCTGCA GGAGCATATA     300

CTGAGATTAC ACCTAAAGAA ATAACTGTTA AAATAATCAT AATTAAAAAG TTAATATGAA     360

AATTTACTAT TACTAAAGGT AAAAGTATAA ATAGTATAAT ACTTTCTACA TAACACCAAA     420

AAGAAGAAGG TGCATGTGCa CCATGTGCAT GtCTTCTTAT TAAATAAAAT GTTAAATTCG     480

TAATTAACGT AAACAGAAAA ATGTTTAAAA TATAGGCAAT AGTATACATA ACAATTAATT     540

TACCTATATT TTTAGCTAAG ACCTGCATCC CTAATCGTAC TTGCAAAAAT TGAATATGAT     600

CTAAGTTATT TCTCTTTTGA AGATACGTGG CAAACTGGTC AATTTTATTA TCAAAATAAT     660

TCAATTTTAC ACCACTCTCC TCACTGTCAT TATACGATTT AGTACAATCT TTTATCATTA     720

TATTGCCTAA CTGTAGGAAA TAAATACTTA ACTGTTAAAT GTAATTTGTA TTTAATATTT     780

TAACATAAAA AAATTTACAG TTAAGAATAA AAAACGACTA GTTAAGAAAA ATTGGAAAAT     840

AAATGCTTTT AGCATGTTTT AATATAACTA GATCACAGAG ATGTGATGGA AAATAGTTGA     900

TGAGTTGTTT AATTTTAAGA ATTTTTATCT TAATTAAGGA AGGAGTGATT TCAATGGCAC     960

AAGATATCAT TTCAACAATC GGTGACTTAG TAAAATGGAT TATCGACACA GTGAACAAAT    1020

TCACTAAAAA ATAAGATGAA TAATTAATTA CTTTCATTGT AAATTTGTTA TCTTCGTATA    1080

GTACTAAAAG TATGAGTTAT TAAGCCATCC CAACTTAATA ACCATGTAAA ATTAGCAAGT    1140

GAGTAACATT TGCTAGTAGA GTTAGTTTCC TTGGACTCAG TGCTATGTAT TTTTCTTAAT    1200

TATCATTACA GATAATTATT TCTAGCATGT AAGCTATCGT AAACAACATC GATTTATCAT    1260

TATTTGATAA ATAAAATTTT TTTCATAATT AATAACATCC CCAAAAATAG ATTGAAAAAA    1320

TAACTGTAAA ACATTCCCTT AATAATAAGT ATGGTCGTGA GCCCCTCCCA AGCTCGCGGC    1380

CTTTTTTGTA ATGAAGAAGG GATGAGTTAA TCATCATTAT GAGACCCGCC GTTAAAATAT    1440
```

```
TCATTTGCAA AGGGCGAAAT GGGTTCTTAC TGAGTTATCT ATTATAAAAA AATAAACATA    1560

GACTTATGAA AAATCTCTCA TAAATCTATG TTTAGTCATG aCATGTGTTA AATATTATTT    1620

CGGGCGCTTC TTATTTATAC AAATCTAATT TAATACTTTT AAATACAGGT ATATTTTCgC    1680

GTTGCTGTTC TACTTCATTT AAGTTTAAAT CTACAGTCAA AATATCTGCG GATTCATTTA    1740

ATTCTCCAAC TAAATCTCCA TTTGGGTTTA TAACTATCGA ATGACCAGCA TATTCTGTGT    1800

TACCATCGAA TCCAGTGCTA TTAGTTCCAA TGACAAACAT ATTATTTTCA ATTGCACGTG    1860

CCTTTAGTAA TGAATGCCAA TGTTGAAGAC GTGACATAGG CCATTGCGCC ACATAAAATG    1920

CAATTTTAGC ACCACTACGA GCAGGATATC TTAATAATTC TGGAAAACGT AAATCATAAC    1980

AGATAAGTTG GGTCACATAA GTACCGTCAG ACAATTGAAA GGGTTCAGCT ACGTATTCGC    2040

CAGCGGTTAA AAATTCATGC TCTCTTAACA TAGGAACTAA ATGAACTTTG TCGTATTCaT    2100

TAATCAGCTG GCCACTTTTA TTCACACTAA AAGCTGTATT AAATATTTGA TTGTTTCTAA    2160

TGTTAGAAAC TGACCCAGCT ACGATATCGA CTTTATATTT TTCAGCTAAA TGTTTAATAA    2220

ATGAAAAACT TTGTCCTAGA TTATTATCTG CTTTTTCATT TAAATGCTCT AAATCATAGC    2280

CATTATTCCA CATTTCAGGT AAAACGACTA CATCTACTTC AGCATTCATA TTTTTTTCGA    2340

ACCATTGCGT TATTTGAGTT TCATTTTTAG AACTATCTCC AAAAACAATC GGTAATTGAT    2400

AAATTTGGAC TTTCATAACA TCACATCCTT GATAGATCTT ATATATAACT TACTAAAAGT    2460

TATGTTGAAA CGCAAAAAAC GAGCACAAGA CATAAAATCA AAGTCCTAGG CTCTACAAAG    2520

TTATATTGAC AGTAGTTGAT GGGGCCCCAA CATAGAGAAA TTGGAACACC AATTTCTACA    2580

GACAATGCAA GTTGGGGTGG GCTCTAACAT AAAGAAATAC TTTTTCTTTA GAAATTAGTA    2640

TTTCTTATAC ATGAGTTTTA CTCATGTATT CCTATTCTTA AGTGCACATT AGCAGCGGCT    2700

AATGTGTAAG AACTACTACA TAATGAATAA CTAATGATTC TTTATCATTT CTGTCCCATT    2760

CCTAACAATA TATTGATTAT TTTTTTATTA CGAAACGATC TTCCACTGGA TTAAATGTTT    2820

TTTCGCCAGC AGCTTCACGA ATATCACCAA ATGGCATTTG AGCAATAAGT TTCCAACTTT    2880

TAGGAATATT AAATTCATTT GAAGTCATCT CATCAACAAG TGGATTATAG TGTTGTAATG    2940

AAGCACCTAT GCCTTTAGTA GCTAATGCAG TCCAAATTGC AAATTGATGC ATGGCATTTG    3000

TTTGAGTTGA CCATATTGCA AAATTATCAT AGTAGTTTGG CATTTGTTCT TGTAAACCAC    3060

TTACAACATC TTGATCTTCA TAAAACAAAA TTGTACCGTA TGAATGTTTG AAGTTATCAA    3120

TTTTTTGTTC AGTTGGCTCG AAATCACGAT TCTCTCCCAT GACTTCTTTT AAAATTGCTT    3180

TTGTGTTATC CCAAAATTTA TTATTGTTGT CATTTAACAA GAGAACAATT CTAGTTGATT    3240
```

```
CATCGCTAAT TGATATCGAA TCTTTCAAAT TATATATTGA ACGTCTTTCT TCCATTGCAT        3360

TGTCAAAAGT CATTGCTTTT TTATCTTTTT TAAATAAGCC CATAATTATT GCTCCTTCTT        3420

TAGTAAAGAA TACTTAATAG ACTAAGTATA AAATTTATAC TCGTACTTGT AAAGCAATAT        3480

TTACGAAAAT TTCAAGAATA TTAATATTCA TTTTCAAATT CCAAATATAA ATGCATTTTC        3540

AACGCATATT TATTATACTT AGATTAATAC TTACATGAAA AAGGGAGGTG TCTCGTGAAA        3600

TGTCATATCA TTGGTTTAAG AAAATGTTAC TTTCAACAAG TATTTTAATT TTAAGTAGTA        3660

GTAGTTTAGG GCTTGCAACG CACACAGTTG AAGCAAAGGA TAACTTAAAT GGAGAAAAAC        3720

CAACTACTAA TTTGAATCAT AATATAACTT CACCATCAGT AAATAGTGAA ATGAATAATA        3780

ATGAGACTGG GACACCTCAC GAATCAAATC AAACGGGTAA TGAAGGAACA GGTTCGAATA        3840

GTCGTGATGC TAATCCTGAT TCGAATAATG TGAAGCCAGA CTCAAACAAC CAAAACCCAA        3900

GTACAGATTC AAAACCAGAC CCAAATAACC AAAACTCAAG TCCGAATCCT AAACCAGATC        3960

CAGATAACCC GAAACCAAAA CCGGATCCAA AACCAGACCC AGATAAACCA AAGCCAAATC        4020

CGGATCCAAA ACCAGATCCA GATAACCCGA AACCAAATCC AGATCCAAAA CCAGACCCAG        4080

ATAAACCAAA GCCAAATCCG GATCCAAAAC CAGATCCAGA TAAACCAAAG CCAAATCCGA        4140

ATCCAAAACC AGACCCTAAT AAGCCAAATC CTAACCCGTC ACCAGATCCC GATCAACCTG        4200

GGGATTCCAA TCATTCTGGT GGCTCGAAAA ATGGGGGGAC ATGGAACCCA AATGCTTCAG        4260

ATGGATCTAA TCAAGGTCAA TGGCAACCAA ATGGGAATCA AGGAAACTCA CAAAATCCTA        4320

CTGGTAATGA TTTTGTATCC CAACGATTTT TAGCCTTGGC AAATGGGGCT TACAAGTATA        4380

ATCCGTATAT TTTAAATCAA ATTAATAAGT TGGGCAAAGA TTATGGAGAA GTTACTGATG        4440

AAGACATTTA TAATATTATT CGAAAACAAa ATTTCAGCGG AAATGCATAT TTAAATGGAT        4500

TACAACAGCA ATCGAATTAC TTTAGATTCC aATATTTCAA TCCATTGAAA TCAGAAAGGT        4560

ACTATCGTAA TTTAGATGAA CAAGTACTCG CATTAATTAC TGGTGAAATT GGATCAATGC        4620

CAGATTTGAA AAAGCCCGAA GATAAGCCGG ATTCAAAACA ACGCTCATTT GAACCGCATG        4680

AAAAAGACGA TTTTACAGTA GTTAAAAAAC AAGAAGATAA TAAGAAAAGT GCGTCAACTG        4740

CATATAGTAA AAGTTGGCTA GCAATTGTAT GTTCTATGAT GGTGGTATTT TCAATCATGC        4800

TATTCTTATT TGTAAAGCGA AATAAAAAGA AAAATAAAAA CGAATCACAG CGACGATAAT        4860

CCGTGTGTGA TTCGTTTTTT TTATTATGGA ATAAAAATGT GATATATAAA ATTCGCTTGT        4920

TCCGTGGCTT TTTTCAAAGC CTCAGGATTA AGTAATTGGA ATATAACGAC AAATCCGTTT        4980

TGTAACATAT GGATAATAAT TGGAACAGCA AGCCGTTTTG TCCAAACATA TGCTAATGAA        5040
```

```
AATATTAATG AACTTACTGT TGTAGCAATA ATAAATGCCA CGATACGATT ACCTTTAATC        5160

GCATTAAATA ATTCTCCAAA GATTACTTTT CTGAATACAT ATTCTTCTAA TAAAGGACCA        5220

ATAATAGATA CAAAGAAGAT AAATATAGGT ATTTTTCGAG CAATAATAAT TAGCTTTTCT        5280

GTATTAGGAC TTACTTGTTG TCCACCATAA ATTTGCGTTA ATACAATGCT CACTACCATT        5340

TGATAAATCA TTACCAATGC AAATCCAAGC AATGCCCATG GAATGATATA TTTTTTAGGT        5400

TCTTTAACTT CTAATTCTAA TTTTGTTGGA TTTTTAATTT TTAAATTAAT TAAAATAATC        5460

GTCGTGGCGG CGATTAAAAA TAGAACAAGT TGTATGTAAA TGACTGCTTT AGTCAGTTCT        5520

ATGCCACTAT ATTGTACAAA TGGTAATTTT TTTACAATGA GAAGCGGTAA AAATTGAGAC        5580

AATATATAAA TAATAACAGT TAGCAATGAT GCCCATAATC tTGTCATAAT TTTCCTCCAA        5640

ATATTTGTTT ATAATTTATT TTATCGTAAA TAACTTGAAG TTACAAAACT TAATTAAAAG        5700

GTTATGACTT GAAATTTTGA CCAAATTTGA TTATTATAAA TGTATGTTAG CACTCTTTAA        5760

TGTTAAGTGC TAAACTTTAG GTTTTTTAAG GAGGAACAAT CATGCTAAAA CCAATTGGAA        5820

ATCGTGTGAT TATTGAGAAA AAAGAACAAG AACAAACAAC TAAAAGTGGn ATTGTTTAAC        5880

TGATAGTGCT AAAGA                                                        5895
```

(2) INFORMATION FOR SEQ ID NO: 2:

     (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6796 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
TTTGAAAAAA CAAGGTACGA TTGGTTTAAT AACATATATG AGAACCGATT CTACACGTAT          60

TTCaGATACT GCCAAAGTTG AAGCAAAACA GTATATAACT GATAAATACG GTGAATCTTA         120

CACTTCTAAA CGTAAAGCAT CAGGGAAACA AGGTGACCaA GATGCCCATG AGGCTATTAG         180

ACCTTCAAGT ACTATGCGTA CGCCAGATGA TATGAAGTCA TTTTTGACGA AAGACCAATA         240

CCGATTATAC AAATTAATTT GGGAACGATT TGTTGCTAGT CAAATGGCTC CAGCAATACT         300

TGATACAGTC TCATTAGACA TAACACAAGG TGACATTAAA TTTAGAGCGA ATGGTCAAAC         360

AATCAAGTTT AAAGGATTTA TGCACACTTTA TGTAGAAACT AAAGATGATA GTGATAGCGA        420

AAAGGAAAAT AAACTGCCTA AATTAGAGCA AGGTGATAAA GTCACAGCAA CTCAAATTGA         480

ACCAGCTCAA CACTATACAC AACCACCTCC AAGATATACT GAGGCGAGAT TAGTAAAAAC         540
```

```
AAAGCGTAAC TATGTCAAAT TAGAAAGTAA GCGTTTTGTT CCTACTGAGT TGGGAGAAAT     660

AGTTCATGAA CAAGTGAAAG AATACTTCCC AGAGATTATT GATGTGGAAT TCACAGTGAA     720

TATGGAAACG TTACTTGATA AGATTGCAGA AGGCGACATT ACATGGAGGA AAGTAATCGA     780

CGGTTTCTTT AGTAGCTTTA AACAAGATGT TGAACGTGCT GAAGAAGAGA TGGAAAAGAT     840

TGAAATCAAA GATGAGCCAG CCGGTGAAGA CTGTGAAATT TGTGGTTCTC CTATGGTTAT     900

AAAAATGGGA CGCTATGGTA AGTTCATGGC TTGCTCAAAC TTCCCGGATT GTCGTAATAC     960

AAAAGCGATA GTTAAGTCTA TTGGTGTTAA ATGTCCAAAA TGTAATGaTG GTGACGTCGT    1020

AGAAAGAAAA TCTAAAAAGA ATCGTGTCTT TTATGGATGT TCGAAATATC CTGAATGCGA    1080

CTTTATCTCT TGGGATAAGC CGATTGGAAG AGATTGTCCA AAATGTAACC AATATCTTGT    1140

TGAAAATAAA AAAGGCAAGA CAACACAAGT AATATGTTCA AATTGCGATT ATAAAGAGGC    1200

AGCGCAGAAA TAATATTTTT ATTTCCTAGA TACATTTTAA GATTGTTAAA TAGAATCATT    1260

AGTGAATCTT ATTTTAAAGA TAGTAAAGGA TTAATCTAAA TAAGTGCGGA TAATATAAAC    1320

ATAACAACAT AATTAAmAGA CATAAATGAC aATAAAAGGA GTATAGAAAT GACTCAAACT    1380

GTAAATGTAA TAGGTGCTGG TCTTGCCGGT TCAGAAGCGG CATATCAATT AGCTGAAAGA    1440

GGAATTAAAG TTAATCTAAT AGAGATGAGA CCTGTTAAAC AAACACCAGC GCACCATACT    1500

GATAAATTTG CGGAACTTGT ATGTTCCAAT TCATTACGCG GAAATGCTTT AACTAATGGT    1560

GTGGGTGTTT TAAAAGAAGA AATGAGAAGA TTGAATTCTA TAATTATTGA AGCGGCTGAT    1620

AAGGCACGAG TTCCAGCTGG TGGTGCATTA GCAGTTGATA GACACGATTT TTCAGGTTAT    1680

ATTACTGAAA CACTTAAAAA TCATGAAAAT ATCACAGTTA TTAATGAAGA AATTAATGCC    1740

ATTCCAGATG GATACACAAT TATCGCAACA GGACCACTTA CTACAGAAAC CCTTGCGCAA    1800

GAAATAGTGG ACATTACTGG TAAAGATCAA CTTTATTTCT ATGATGCGGC TGCTCCAATT    1860

ATTGAAAAAG AATCTATTGA TATGGATAAA GTTTACTTAA AGTCCCGTTA TGATAAAGGT    1920

GAAGCTGCAT ATTTAAACTG TCCTATGACT GAGGATGAAT TTAATCGCTT TTATGATGCA    1980

GTATTAGAAG CTGAAGTTGC GCCTGTAAAT TCATTTGAAA AGAAAAATA TTTCGAGGGT    2040

TGTATGCCTT TTGAAGTAAT GGCAGAACGC GGACGCAAGA CATTACTATT TGGACCAATG    2100

AAACCAGTAG GATTAGAAGA TCCAAAGACT GGGAAACGTC CTTATCGGT GGTTCAATTA    2160

AGACAAGATG ACGCTGCTGG TACACTCTAC AATATTGTTG GCTTCCAAAC GCATTTAAAA    2220

TGGGGAGCTC AAAAAGAAGT CATTAAATTA ATTCCAGGCT TAGAAAATGT TGATATTGTT    2280

AGATATGGTG TGATGCATAG AAATACCTTC ATTAATTCAC CGGACGTATT AAACGAGAAA    2340
```

TATGTAGAAA GCGCAGcTAG CGGCTTAGTT GCAGGTATCA ATCTTGCGCA TAAAATATTA   2460

GGCAAGGGTG AGGTAGTATT TCCGAGAGAA ACAATGATTG GAAGTATGGC TTACTATATT   2520

TCTCATGCTA AAAACAATAA GAATTTCCAA CCTATGAATG CTAACTTCGG GTTATTACCA   2580

TCTTTAGAAA CTAGAATTAA AGATAAAAAA GAACGCTATG AAGCACAAGC TAATAGAGCT   2640

TTGGATTACT TAGAAAATTT CAAAAAAACT TTATAAAATA GTTAGAAAGA CTAGATATGC   2700

TATTCATTCT TAAGTCATCA ACGAGTAAGT AATGACTTTC TAAATGGAAA ATACTTATCC   2760

TAGTCTTTTT AATTTTGGAA TTGTTACGTA TTTCTGACAA TTTAGAATTC GCATTCAAAA   2820

AATATCTAAA TAAATAACAC GCAATAAGTT GATTGATGTA ACATGTAAGA GAATGTTTTA   2880

AATAAACTTT ATTTAAAAGG CAATGAAATA ATAAATGGCA AGGCTATTAA TAAAGACTTT   2940

TAGTAATTAA TTTAAAAAAG AGGTATTCTA ATTAACAGGT TTTCCGATTA GTTACAATTA   3000

TTTAATTCTC AAAAGATTTA GAATTGATTA TCAAATTACT GTAAGCCCTT TGCTGTATAT   3060

GCTACAATTC TTATTGATGG AGGGTAAATG TATTGAATCA TATTCAAGAT GCGTTTTTAA   3120

ATACATTGAA AGTTGAACGG AATTTTTCGG AACACACATT GAAATCATAT CAAGATGACT   3180

TAATTCAGTT TAATCAATTT TTAGAACAAG AACATTTAGA GTTGAATACT TTTGAATACA   3240

GAGATGCTAG AAATTATTTG AGCTATTTAT ATTCAAATCA TTTGAAAAGA ACATCTGTTT   3300

CTCGTAAAAT CTCAACGTTA AGAACTTTCT ATGAATATTG GATGACGCTT GATGAGAACA   · 3360

TTATTAATCC ATTTGTTCAA TTAGTACATC CGAAAAAAGA AAAATATCTT CCGCAATTCT   3420

TTTACGAAGA AGAAATGGAA GCGTTATTCA AAACTGTAGA AGAGGACACT TCAAAAAATT   3480

TACGGGATCG AGTTATTCTT GAATTGTTGT ATGCTACAGG CATCCGTGTT TCGGAATTAG   3540

TAAATATTAA AAAACAAGAT ATAGATTTTT ACGCGAATGG TGTTACCGTA TTAGGAAAAG   3600

GGAGCAAAGA GCGCTTTGTA CCGTTTGGTG CTTATTGTAG ACAAAGCATC GAAAATTATT   3660

TAGAACATTT CAAACCAATT CAGTCATGCA ATCATGATTT TCTTATTGTA AATATGAAGG   3720

GTGAAGCAAT CACTGAACGC GGTGTACGAT ATGTTTTAAA TGATATTGTT AAACGAACAG   3780

CAGGCGTAAG TGaGATTCAT CCCCACAAGC TCAGACATAC ATTTGCAACG CATTTATTGA   3840

ATCAAGGTGC AGACCTAAGA ACAGTACAAT CGTTATTAGG TCATGTTAAT TTGTCAACAA   3900

CTGGTAAATA TACACACGTA TCTAACCAAC AATTAAGAAA AGTGTATCTA AATGCACATC   3960

CTCGAGCGAA AAAGGAGAAT GAAACATGAG TAATACAACA TTACATGCAA CAACAATTTA   4020

TGCTGTAAGA CATAATGGGA AAGCAGCTAT GGCTGGAGAT GGGCAAGTAA CGCTTGGTCA   4080

ACAAGTCATC ATGAAACAAA CGGCAAGAAA AGTGCGACGT TTATATGAAG GTAAAGTGTT   4140

```
ATTACAACAG TTTAGTGGTA ACTTAGAAAG AGCTGCTGTT GAATTGGCAC AAGAATGGCG    4260

AGGCGATAAA CAATTACGTC AATTAGAAGC TATGCTAATT GTAATGGATA AAGATGCTAT    4320

TTTAGTTGTC AGTGGAACTG GCGAAGTTAT TGCTCCAGAT GATGACCTTA TCGCTATTGG    4380

ATCAGGAGGC AACTACGCAT TAAGCGCAGG ACGTGCATTG AAACGCCATG CATCGCATTT    4440

GTCTGCTGAA GAAATGGCAT ATGAGAGCTT GAAAGTAGCG GCTGATATTT GTGTCTTTAC    4500

CAACGATAAT ATTGTTGTCG AAACACTATA ATAATCAGAG CACGATAAAT AATTACGAGC    4560

AATTAATTTT AGTTAAAAGA CGGAGGAATG AAATTAATGG ATACAGCTGG AATAAGATTA    4620

ACTCCAAAAG AAATCGTATC TAAATTAAAT GAATACATCG TTGGACAAAA TGATGCTAAA    4680

CGTAAAGTGG CAATTGCCCT ACGTAATCGA TACAGAAGAA GTTTATTAGA TGAGGAATCA    4740

AAGCAAGAAA TTTCACCTAA AAATATTTTG ATGATTGGAC CAACCGGCGT TGGTAAAACT    4800

GAAATTGCAA GAAGAATGGC CAAAGTTGTC GGCGCGCCAT TTATAAAAGT AGAAGCTACT    4860

AAATTTACTG AGGTAGGTTA TGTAGGACGA GATGTTGAAA GTATGGTTAG AGATCTTGTT    4920

GATGTTTCAG TAAGATTAGT CAAGGCGCAG AAAAAATCAT TGGTACAAGA TGAAGCAACA    4980

GCTAAGGCCA ATGAAAAACT TGTTAAGTTA TTAGTTCCAA GTATGAAAAA GAAAGCGTCT    5040

CAAACGAATA ATCCTTTAGA GTCACTTTTC GGAGGTGCAA TTCCAAATTT CGGACAAAAT    5100

AACGAAGATG AAGAAGAACC ACCTACTGAG GAAATTAAAA CAAAACGTTC TGAAATTAAG    5160

AGACAGCTAG AAGAAGGCAA ACTTGAAAAA GAAAAGGTAA GAATTAAAGT CGAACAAGAT    5220

CCTGGTGCTT TAGGTATGCT AGGTACAAAT CAAAATCAGC AAATGCAAGA GATGATGAAT    5280

CAATTAATGC CTAAAAAGAA AGTTGAGCGA GAAGTTGCTG TTGAGACGGC AAGGAAAATC    5340

TTAGCTGATA GTTATGCGGA TGAACTAATT GATCAAGAAA GCGCTAACCA AGAAGCGCTT    5400

GAATTAGCAG AACAAATGGG TATCATCTTT ATAGATGAAA TCGACAAAGT TGCGACGAAT    5460

AATCATAATA GTGGTCAAGA TGTCTCAAGA CAAGGTGTTC AAAGAGATAT TTTACCTATA    5520

CTTGAAGGTA GCGTTATTCA AACCAAATAT GGTACTGTGA ATACTGAACA TATGCTGTTT    5580

ATAGGTGCTG GAGCTTTCCA TGTATCTAAG CCGAGTGACT TGATACCAGA ATTGCAAGGT    5640

CGTTTTCCGA TTAGAGTTGA ACTTGATAGT TTATCGGTAG AAGATTTTGT AAGAATTTTG    5700

ACAGAACCAA AATTGTCATT AATTAAACAA TATGAAGCAT TGCTTCAAAC AGAAGAAGTT    5760

ACTGTAAACT TTACCGATGA AGCAATTACT CGCTTAGCTG AGATTGCTTA TCAAGTAAAT    5820

CAAGATACAG ACAACATTGG TGCACGTCGA CTTCATACAA TTTTAGAAAA GATGCTAGAA    5880

GATTTATCAT TCGAAGCACC AAGTATGCCG AATGCAGTTG TAGATATTAC CCCACAATAT    5940
```

```
AAATATACAA AAGGAGAAAA ATTCATGAGC TTATTATCTA AAACGAGAGA GTTAAACACG        6060

TTACTTCAAA AACACAAAGG TATTGCGGTT GATTTTAAAG ATGTAGCACA AACGATTAGT        6120

AGCGTAACTG TAACAAATGT ATTTATTGTA TCGCGTCGAG GTAAAATTTT AGGATCGAGT        6180

CTAAATGAAT TATTAAAAAG TCAAGAATT ATTCAAATGT TGGAAGAAAG ACATATTCCA         6240

AGTGAATATA CAGAACGATT AATGGAAGTT AAACAAACAG AATCAAATAT TGATATCGAC        6300

AATGTATTAA CAGTATTCCC ACCTGAAAAC AGAGAATTAT TCATAGATAG TCGTACAACT        6360

ATCTTCCCAA TTTTAGGTGG AGGGGAAAGA TTAGGTACAT TAGTACTTGG TCnAGTACAT        6420

GATGATTTTA ATGaAAATGA TTTGGTACTA GGTGAATATG CTGCTACAGT TATTGGTATG        6480

GAAaTCTTAC GTGAGAAGCA TAGTGAAGTA GAAAnAGAAG CGCGCGATAA AGCTGCTATT        6540

ACAATGGCAA TTAATTCATT ATCTTATTCT GAAAAAGAAG CGATTGAACA TATCTTTGAA        6600

GAACTTGGCG GTACGGAAGG CCTATTAATC GCATCAAAAG TTGCAGATAG AGTTGGTATT        6660

ACTAGATCTG TAATTGTAAA TGCACTACGT AAATTAGAAA GTGCTGGTGT AATTGAATCA        6720

CGTTCTTTAG GAATGAAAGG TACTTTCATT AAAGTTAAAA AAGAAAAATT CTTAGATGAA        6780

TTAGAAAAAA GTAAAT                                                        6796
```

(2) INFORMATION FOR SEQ ID NO: 3:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 2073 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
ATCCTAAAAT TnAAAATTAT CACGCCTTTT GaACAGCTTT GTAACCaTCt GGACGATCAT        60

kAAATTCCaA TGTAAATCCT GGTTTAAaGT TGATCTTTAA CCTTATTTAA AyCACCAATT        120

GTACGTATAT TATGTTGTTT AGCAAAATCA CGTTTTACAG CTAAAGCATA CGTATTGTTA        180

TACTTCATTG GTTTTAACAT AGTCATTTGA TATTTCTTTT CAAGACTTTG CTTAGCTTGT        240

TCATAAACTT TTTTCTCTTC TTTTGACTTC AATGGTTCTT TTGTTAATTC ACCTAAAACT        300

GTTCCAGTAA ATTCTAAATA CCCATCTATA TCGTCAGATT TTAAAGCATT AAATAAAAAT        360

GCTGTTTTGC CCATACCATC TTTCACTTCT ACAGTATTTT TGGTCTCTTC TTCTATTAAA        420

ATTTTATACA TATTTGTAAT AATCGATGGC TCGGAGCCAA GCTTTCCAGC TAACGTAATT        480

TTATCACCTT TTTGTGCAAA CATAGGAATA GCGATAGCCA GTATAATAAT CATCACTATA        540
```

224

```
TCAAATATAA TTGCCAATAA GGCTGCTGGA ATTGCACCTA ATAATATCAA CGATGCATTG        660

TTACGGTCTA TACCTAATAA AATTAAATCT CCTAGTCCGC CTGCACCAAT TAATGCTGCT        720

AGTGTTGCTG TACCTATAAT TAATACCATA GCCGTTCTTA CACCAGCCAT TATAACAGGC        780

ATTGCTATCG GAAGTTCGAC TTTAGTTAAA CGTCTAAATG GTTTCATACC TATACCTTTA        840

GCCGCTTCAA TGAGTGATGG ATCAACTTCT TTAATTCCAG TATACGTATT CCTTAAAATT        900

GGTAACAACG CATACACTAC AAGTGCAATA ATTGCTGGCA CACGACCGAT ACCAAATAAA        960

GGAATCATTA AACCTAATAA TGCCAACGAT GGTATGGTTT GAAGAATTGC CGCAATATTC       1020

ATTACGATTT CAGATATCGT TTTAGTCTTC GTTAATAAAA TACCTAATGG TACCGCAATA       1080

GCAGTTGCAA TCAATAATGC GATAAATGAT ATTTGAATAT GTTCTATCAT TGTCGAAAAG       1140

AGTTGCCCCT TACGTTCACT CAATATGTCg AAAAAGTTAG TCATGTTGAG CTACCTCCTT       1200

TTTCTGGGAC AAATATTTGA AGATATCTTT CCTATCAATA ACATATTGAC CTACGCTATC       1260

TTCTTGCATG ACAATGACAC GCTCGCTCTC TGATAAAAGT TGATACAATA CTTCAATTGG       1320

TTGATTGTCA TAAACAATTG GATAAGCGCT CATAGATGTA ACCTCATCGA TTGGTTTCAT       1380

AATATCCAAG TCACGGATAA TTGCGTTCTC TTCAACACAT GGCGCATCAT CTTCTAAATG       1440

ACTACCCATA AATTGTTTAA CAAATTCACT TTGAGGATTA TTTTTAAATC CTTCTGGTGT       1500

GTCAATTTGT TCAATATGCC CTTCATTCAA AAGACAAATC TTATCACCAA GTTTCATCGC       1560

CTCTTGAATA TCATGTGTAA CAAATATGAT TGTCTTCTTA ATTTTAGTTT GTAATTCAAT       1620

TAAATCATCT TGAAGTTTTT CTCGGCTGAT TGGGTCTAAT GCACTAAACG GTTCATCCAT       1680

TAAAATAACT GGTGGATCAG CTGCTAACGC ACGTATAACT CCTACACGTT GTCGTTGCCC       1740

CCCTGACAAT TCATCAGGTT TTCTGTTTTT ATATTTTTCA GGTTCTAATC CAACCATTTC       1800

AAGTAATTCA TCTACTCTTT TATCTATATC TTTTTCTTTC CACTTTTTCA TTTGTGGCAC       1860

TTGTGCAAtA TTTTCTTTGa wTGTCaTATG TGGGAATAAT GCAATCTGCT GcAATACGTA       1920

TCCAATATCC CAACkCATTT CGTATACTGG ATAATCACTT ATTGGTTTAT CTTTAAAATA       1980

AATATAACCT TCACTTAAGT GAATGAGTCG ATTAATCATT TTTAATGTCG TAGTTTTTCC       2040

ACAACCTGAA GGTCCAATTA GCACAAAAAA TTC                                    2073
```

(2) INFORMATION FOR SEQ ID NO: 4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 13321 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
ACTATTCTAG CTTCATCAGT TATCATATAT TCTTTGAAAC ACTTGTAAGA AAATATAATG      60

AGTATTTACT ACATAATGAT ATTTCAAATT AGAAAAAAGG AAGTTATGAT TTAATGGCCT     120

TGAGCCTATC ATAACTTCCT TTTATCATTT TATTGTTGTG TTGATGTTTC GATAACGTGG     180

TACATCTTAT CAAACATCAA TTCGAAACCA TGCACCATGG CATCATGATA TTCTTTTTTC     240

TTTTGCTTGT ATTCTAAATT AGTAAATCGT CTTTCTTTTT CAACTAATGA ACGATAATAA     300

AATAGCATTT GGGTGCCACC TGTTTCACGT TCAAAAAATT CTACCTCAAT GACATCTTGC     360

GTTTCACTTA GTCCAGGCAT ACCGATAGTC ATCTTAACGT ATTCATCCAT AACTAAAGAT     420

TCATAAATGC CTTCAATCAC ATTTACTTTG CCATTACGTT GTTGATCTAC AATACGATAT     480

TTACCGCCTT CTTTAACGTC CGCTTCAATC TCTTTATTCG TTCTGGCTGA TGTCATAAAC     540

CATTGTTTCA ACAAATCTTT CTTTGTCCAA GCTTCGTATA CTAACTCTGG AGAAAATTTA     600

TAAAGCTTTT CAATTTCAAC TTCGACATGT TCATTCTCTA CATTAAATTT TGCCACTGTT     660

GTCCACCCAC TTTCGCTCTT ACTTTTATTT TAACGTATTT TTGCTCAGTT CCAAACATAG     720

ATGATCATCA TTTTTAAAAG ATTAGCGTTA TACGGTGAGT ACAACATGAT CTGTTAATAT     780

AACAAGCCAC CTTACTTGGC TACATCGATA TATTGTTAAG CATTAATGTT TCATTTCTTG     840

ACTAGTGTTC TTTTTTAGCT TTGGAAAATT AAATAAAATC GCAATAAGTC CGCATACACC     900

TAATAATATA GGATAAATGC TGTATGGGAA TAACATTAAC GGTGAAATAC CAGCTACACC     960

AGCCGCTGaA ATGACTTGCG GGCTATATGG TAATAAACCT TGGAAGCAGC CTCCAAATAT    1020

ATCAAGAATA CTTGCTGATT TCCTTGAATC TACATCATAT TCATCTGCAA TATTTTTAGC    1080

TAAAGGACCT GACATAATAA TAGAGATGGT GTTGTTTGCC GTGGCAATAT CTGCGACACT    1140

TACCAAACTA GCAATTCCTA ATTCTGCGCC ACGCTTTGAT TTCACTTTAG AGCGAACAAA    1200

TTGCAACAAC CATTCAATAC CACCATTGTG TTGAATAATA CCGACTAAAC CACCAATTAG    1260

CAACGCAATC ATAGCAATAT CTTCCATGCT TATAATACCT TTGGACACTG CATCTAGTAG    1320

CCCCATCCAA CCGAATGAAC CATCTATGAG ACCAATGATT CCGGCTAATA ATGTTCCGCC    1380

AATCAATACG ATAATGACAT TTACACCTAA TAATGCTAAT ACCAATACTA AGATATACGG    1440

TACAACTTTA ATTAGATTAT AATCATAGTt TTTAGCATGA TTTAAAGAAA TGCCATTCGT    1500

TAAGAAATAC AGAATAATAA TCGTTAAAAT AGCACCTGGC AATACAATTT TAAAGTTTAC    1560

TCTGAATTTA TCTTTCATTT TCGTATGTTG TGTTCTAACC GCAGCAATTG TTGTATCTGA    1620

AATCATTGAT AGATTATCGC CGAACATTGC ACCTCCAACA ACTGTAGCCa tTGctAGCGC    1680
```

226

```
TCCTACAGAC GTCCCCATAG ATATAGAAAC AAACATACAA ATCACAAACA ATCCTACAAT      1800

AATTAAATTT TCTGGGATTA ATGATAGTCC TAAATTAACT GTCGACTTTA CGCCACCCAT      1860

TTTTTCAGCT GTATTTGAAA ATGCACCTGC TAAAATAAAA ATCAACATCA TTAAAACAAT      1920

GTTTGAATGG CCTGCACCTT TCGTGAAGAC CTCAACTTTT TTAGCAAATG ATTCTTTTCG      1980

ATTCATTAAT AACGCCACAA TTACCGTTAT CGTAATTGCA ACATTTAATG GCATTGAAGT      2040

AAAATCACCT GTGATAATAC CTACGCCTAA AAACAACGCC ACAAATAATA ACAAGGGGAA      2100

TAATGCCCAA GCATTGCTCT TTTTATGTAC TTCCATCCTT TTTACCTGCT TTCCAATTAA      2160

AAATACCTCT TTCTCACAAA CGATGAAGAA AGAGGTTTTC ATGTGCTTTA CCTGCTTATC      2220

TTCAAACCAT TACGGTTACT GGAATTGGCA CATTCGAGAT GTTGCCGAGG CTTCATAGGG      2280

CCAGTCCCTC CACCTCTCTA GATAAGTGAT GCTTATTTAC GTTTACGTTA CAAGATAATC      2340

CTTAGTACGT CAATCATAAA TTAATCAGGA GTCGTATAAT ATTTTTCATA AACAATCATT      2400

GCTACTGTAA TAATAATCAA AACAATAATG CTAATAACAA GTAAAGCCA CCATTTAAGC        2460

ATTAATGCAA TAAAAATGAA CACGATAGAC ACACTTACTA ATATTAATGA TATGACTTTA      2520

AATTGCTGAA CACGTTGCTT GGAGATGACT TTCAACTGTT TGTTTGATAG ACGCGTATTT      2580

TTTATACTGA TTCCCAGTAT ATTTTCTAAT ATTTGAACCA ATACGATACT TATTGCAAAT      2640

ATAATAATTG GTAAAACATC ATAGCTCCCT ATAGTTAATG TATAAATTAC AAATCCAATG      2700

TAAAGTAACC CTGAGACAAA GGATAAAAAG TATGCGACGT ATTTGTTAAA CTTAATGATA      2760

TGCTTTTTAA CGTTTTGATG TGTAAACCAT ACATTCGAAA CGATCGCAAC TGCTACAAAT      2820

AATGTGAATA CTATATATAA TGGTAATTTT TGTTCAGGAA AAACAGTCGC TATTCCAAAA      2880

GCTAATGCTA AAATCAAAAA TAATATAGCT CTAGATACTA TTAATGCCAT AATAACAACC      2940

CCTTTGTTTA ATATCGAGTT TGCAAATTTA CGTTTATCAG CGTTTCTATG ATCAGTACTT      3000

CTACGGGTAG CGTTTCTATG TAATTTACAT CATCTTAACA TATAAATACT TCGCTATTTA      3060

ATTGAAAACA TATCCTATTA TTCTTTGTCC GTTCTGACGT TTAATATCTA GCCTTAGGCA      3120

TTTCACTTGT TAATGAATTT AACTTTCTTC CACTAACCGT CCCTAAACCC AATCCCGCAA      3180

CAGTTTTTAA CTTTTTCGTT GTTGTCCTGA CATCCTCATT AAGAAAGTTT ATTCTGCTTA      3240

AAACTTATAA TCCACACCCT GAGCAAACGC TCCTTATGAC AGAGTATTAA AATAAGCCGA      3300

TAAAGATACA CACCTTTACC GACTATTTAA AATACACTTC ACCAATTCAT TTTAATTTAA      3360

TGGATTGAAG TAACTAAATT AATATTATGT TGTTCAATTA AAAGCTTCAT ACAAACCTAA      3420

TCTATTTGCA CTCCACCGCT AACACCGAAC ACTTGTCCGG TTGTATAACT TGATTCTTCT      3480
```

```
GTTTTTTGAC CAAATGTTGG GATTTTACTT TGAGGTTGTC CACCAGAAAT TTGTAATGGT    3600

GACCAGAATG GACCAGGCGC TACACAGTTC ACTCTAATTC CTTTTGGTCC TAATTCTTCT    3660

GAAAAACTTT TAGTTAATGA AATAATTGCT GCTTTTGAAG CGGCATAATC ATGAAGAATA    3720

GGACTAGGAT TATAACCTTG TACAGATGAT GTCGTTGTAA TTGACGCACC CGGTTTTAAA    3780

TATTCCAATG CTTTTTGAAC TGTCCAAAAT AGCGGATAGA CATTCGTTTC AAATGTTTCT    3840

GTAAATGCCT CAGTTGTAAA TCCATGAATA TCATCATGAT ACTGTTGATG TCCAGCAACT    3900

AAAGTAACAT TATCTAAGCC ACCTAATTGT TGATATGCTT GTTCAACAAG GTCATAGTTG    3960

AACTGTTCAT CTCTTATATC ACCAGGAATT AACACTGCCT TTTGACCACT TTCTTCAATC    4020

ACTTGGCGTA CTTCTTGTGC ATCTTGTTCT TCACTCGGAA GATAGTTAAT CGCTACATCT    4080

GCACCTTCTT TAGCATACGC AATTGCTGCT GCACGCCCTA TTGCTGAGTC ACCACCTGTG    4140

ACTAATATTT TATAGCCTTG TAAGCGTTGA TGACCTTGGT AAGACGTTTC GCCACAATCG    4200

GGTGCTGGCG TCATTTCAGA TTGTAAACCC GGTACCTCTT GTTCTTGTTT TTCATAATCC    4260

GTTGTTTTAA ATTTTGTTCT AGGATCTTGA GCTGCCATTT TTTTACATCT CCTTATTCGC    4320

TTAATGGTTA TTATTTACCC AATCTTCCTA GGAACTTAAT CATGATTACA CTAAAAATTA    4380

CTTTCTTCTT TATAAAAACA AGCTCGAATT ATTCATGCAA TAGTCTCTTT ACAAATTCAA    4440

CAAAATACTC AGGTACTTTT TCCAGAATCC TTTCATCCGG TTTATATTGA GGATGATGTA    4500

AATCATATTC ACTATGAGAA CCAATTAACG CAAATACACT TGGAAAATGT TGACTATAAC    4560

CTGAAAAATC TTCTCCAATC GTAAGCGGCT GTTCCATCAT TCCCACCTTA TATCCAACAT    4620

GTTGGGCTAC TGCAATTGCT TTATGCGTCA ATGCCTCATC ATTCATCACA GCGCCAGGTA    4680

AATGCGTATA ATTTAAATTA ATTTTCATAT TATATGCTTG AGCCAATCCG TCCGCAATAT    4740

CTTGTAATCG TGTTTCTACA AGCTTTCGTA CCACAGGATC AAAACTACGC ACTGTGCCTT    4800

GTACATACGC ATGATCAGCA ATGACATTCC AAGTATTACC ACATGATATT TGTCCAATTG    4860

TTACTACCGC TTCATCAAAC GCAGATAGAT TTCTACTAAC TATGGATTGA ATACTATTAA    4920

TCAATTGCGC CAACACAATA ACTGGATCGT TGCATTGTTC TGGCTTTGCA GCATGACCAC    4980

CCACGCCTTT AATATGAAAC TCAAAACGAT CTACTGCTGA TGTAATTGCC CCTGTTTTGA    5040

TTGCAAATGT ACCTACCGAA CGCGATGGGT CATTATGAAA ACCCAATACT GCTTGTACAT    5100

CTTTTAATGC ATGTGTTTCA ATAATTTTAA AAGCGCCATG TCCTAGTTCT TCTGCTGATT    5160

GAAAAATGAA TTTAACACGC CCAGTAAGAG TGCCCTCAAT TTCTTTTAAT TTTACAGCTG    5220

TAGCCAAAAT ACTAGCCATG TGAATATCAT GACCACACGC ATGCATAACA CCTTCATTTT    5280
```

```
CAGCTATACA ACTCAGACCT TGTCCCACTT CAGCAACAAG CCCAGTCGCA AGTGGTAAGT    5400

CTAATATTCT AATATGATGT TCTGTTAAAA TATCTTTAAT TTTTTGTGTA GTCTTAAATT    5460

CTTTATCGGA TAGTTCTGGA AATTGATGAA AATACCTTCT CCAGGTAACA GCTTGATCTT    5520

TTAATCCCAT CGGTCATTCC CCTTCCTTAA GTCAATGATA TGTTGTCTAC CCTACGATGA    5580

TCATCTTTGA CTATTAAACG ATGATTTCAC AACAATGTAC TCTTGTTAAT TGCTTTCGTT    5640

AATGATAGAC AGTTGTTTAA TAATATCGTA ACACTGTTGT CAAACTATTC TAACTTTTAT    5700

AATTGAGACT CTATACAAAA ACGTGTTCTC GAATATACTT GTTTTTACAA ACCACAAAAA    5760

GCTCTAAACA TTAGTTTAAA CCAATGCTTA GAGCTTTCTA ATTATTTTAT GCTTTAAAAG    5820

ATACTGTGTT ATCTACGATG ACCTTACCGT CTTTAATAAC TTTTTCTGCG TGATTGATAC    5880

CAAAATGATA TGGAATATAT TCATGATTTG GTGCATCCCA AATTACTAAA TTAGCCTTAT    5940

CACCTGTGTT AATTGTACCC GCGTTAATGT CTATTGCTTT AGCAGCATTG ACCGTAACAG    6000

CATTCCAAAC TTCATTAGGT GATAGCTTTA ATTTCAAGGC TGCAATCGCC ATAACAAGTT    6060

GTAAGTTGTT TGTGACACTA CTACCAGGGT TATAATCAGT TGCTAATGCA ATCGCACCGT    6120

TATTGTCAAG CATGCCTCTT GCATCTGCAT AATCTTCTTT ACCTAAATAG AACGTCGTTG    6180

CAGGTAAGAG GACAGCTACA GTATCACTAT TTCGCAACTT TTCTTTTCCT TTATCACTAG    6240

AAGCTACTAA GTGGTCTGCT GATATTGCTT GTTCATCAAT TGCTAATTCC AGTCCGCCTA    6300

ACGGATCAAT TTCATCCGCA TGTATTTTCA CTTTAAAACC TGCTTCTTTG GCTTTTTGCA    6360

TATAATGTTG CGATTGTTCT ATTGTAAATA CACCTGTTTC ACAGAAAATA TCCGCAAAGT    6420

CTGCATATTG TTTTACTTCC GGAAGTAACG CAATCATTTC TTCTAAAAAT GCCTCATTTG    6480

AACTTGCCTC TTTAGGTACA GCATGAGGCC CTAGGAAAGT ATGTTTCATG TCTAAATCAT    6540

ATTTCTCAGC TAAACGATTA GACACTTTCA ATTGCTTCAG TTCATTTTCT CTATCTAATC    6600

CATAACCACT CTTACTTTCA ACTGCAAGCA CGCCGTGTTT AATCATAGTA AGCAAATCAT    6660

GCTCTGCTTT TTTAAACAAG TCATCTTCGG ATGTTTCTCT AGTAGCATTA ACGGTAGATA    6720

ATATGCCACC ACCCATTTCT AATATTTCAA GGTAAGACTT ACCTTGACGT TTTAATGACA    6780

TCTCATGTTC TCGAGATCCA CCAAATGTTA AATGGGTATG TGCATCTACT AATGCTGGGG    6840

ACACTACCTT CCCACTAGCA TCAATCGTCT CAGTCGCATC GTAGTCATCT GTATGTGTTC    6900

CAGCATATAC AATTTTGCCA TCTTTAATGA CAACTGTACC ATTTTTCACA ACATTTAATT    6960

CATCTAATTC CTTACCCTTC AAAGGTTTAT CTGTTGATCT CGGTAAAATT AATTCTGCTA    7020

TATGATTAAT TATTAAATCA TTCATTACTT ATCACCTGCT TTATCAATCA TTGGAATATG    7080
```

229

```
AACACCCATA CCTGGGTCAG TCGTCAATAC ACGTTCCAAT CTTCTTTCAG CACGCTCTGA    7200

TCCATCTGCT ACAACAACCA TACCCGCATG AAGTGAATAT CCCATGCCAA CACCGCCACC    7260

GTGATGGAAT GAAATCCATG AACCACCTGC AGCTGTGTTA ATGAGTGCAT TCAATACAGC    7320

CCAATCACCA ACCGCGTCAC TACCATCTTT CATACTTTCT GTTTCACGGT TAGGACTAGC    7380

AACTGAACCA GCATCTAAAT GGTCTCGTCC AATAACAATT GGTGCTGAAA TTTCACCGTC    7440

ACGTACAAGA CGATTTAAAG CTAAGCCCAT TTTCGCTCTT TCTCCATAGC CTAACCAAGC    7500

AATACGTGAT GGTAGTCCTT GATATGAAAT TTTTTCTTCA GCTAAATCAA GCCATCTTAA    7560

TAACTTTTCA TTTTCTGGGA AAAGTTTGCG CATTTCTTCA TCCGCACGCT CGATATCTTT    7620

TGGATCACCA CTCAACGCAG CAAAGCGGAA TGGCCCTTTA CCTTCACAGA ATAATGGTCT    7680

AATGTAAGCT GGTACAAAGC CTGGGAAGTC AAAAGCATTT TTCACTCCGT TATTGAAGGC    7740

TACTTGACGA ATATTGTTAC CATAATCAAA TGCTACAGCG CCACGTTTTT GGAATTCAAG    7800

CATTAATTCA ACATGCTTTG CCATTGAAGC TTGTGACAGT TCAACATATT TTTTCGGATC    7860

TTTTTCACGC AATACTTTCG CTTCTTCTAC AGAGTATCCT TGTGGCACAT ATCCATTTAG    7920

CGGATCATGT GCACTTGTTT GGTCAGTAAT AATGTCAATT TTAAATCCTT TTTCTAGAAT    7980

CGCTTGATGG ATGTCTACAG CATTTCCAAC TAACCCGATT GATAATCCTT CTCCACGTTC    8040

TTTCGCCTCT TCTGCTAATT TTAATGCTTC ATCTAAATCA GCTGTTTTAA CATCACAGTA    8100

TTTCGTATCA ATTCGCTTAT CAACACGTGT TTCATCAACA TCCACGCAAA TTGCTACCCC    8160

ATGATTCATA GTAATTGCTA ACGGTTGCGC ACCACCCATA CCACCTAAAC CTGCTGTCAG    8220

TGTAACAGTG CCTGCTAAAT CTCCATTAAA GTGTTGATTA CCTAGCTCGG CAAATGTCTC    8280

ATAAGTACCT TGCACAATAC CTTGAGAACC AATATATATC CAACTACCGG CTGTCATCTG    8340

TCCATACATG ATTAAACCTT TTTTATCTAA TTCATTAAAA TGATCCCAGT TTGCCCATTC    8400

AGGCACTAAT ACTGAATTTG AAATTAATAC ACGTGGCGCT TCTTCATGTG TTTTAAATAC    8460

AGCAACTGGC TTTCCTGATT GTACTAACAT TGTCTCATCT GATTCTAATT CTCGTAACGT    8520

TTTCTCTATT GCTTCAAAAG CTTCCCAATT ACGTGCTGCT TTTCCAATAC CACCATAAAC    8580

AACTAAATCT TCTGGTCTTT CAGCAACTTC TGGGTCTAAA TTGTTGTATA ACATTCTAAG    8640

TACTGCTTCT TGTTCCCAAC CTTTACACTC AATACTCAAA CCTTTTTTTG CTTGAATTTT    8700

TCTCATAAAA TTCGCTCCTG TTCTTTTAAG AAGTTAATTC CACTAAATTT AAAACGCTTA    8760

CATTATTATC TTCAATATTC ATTATAGTAT GTTAAAATAT AGCCAACAAA TATAAATAAA    8820

CTAATTATCC ATAGCTTGAA TCTATAAATA AAAGGAGCAA AACACATGAA AATTATTCAG    8880
```

```
CATATTAGCC AGCCATCTTT AACTGCTACG ATTAAAAAAA TGGAAGCAGA TTTAGGTTAT    9000

GACTTATTTA CACGTTCAAC AAAAGACATC AAGATTACCG AAAAAGGAAT ACAGTTTTAT    9060

CGTTATGCGA GCGAATTAGT TCAACAATAT CGATCCACGA TGGAAAAAAT GTATGATTTA    9120

AGCGTTACAT CAGAACCAAG GATAAAAATT GGGACTCTTG AATCTACGAA TCAATGGATT    9180

GCGAATTTAA TTCGAAAGCA CCATTCCGAC TACCCTGAAC AGCAATATCG TTTATATGAA    9240

ATACATGATA AACATCAATC TATAGAGCAA TTACTGAATT TTAATATTCA TTTAGCTATA    9300

ACAAATGAAA AAATAACCCA CGAAGATATA AGATCCATTC CTTTATATGA GGAATCTTAC    9360

ATTTTATTAG CACCCAAGGA AACATTTAAA AATCAAAATT GGGTAGATGT TGAAAATTTG    9420

CCACTCATAT TACCAAACAA AAATTCTCAA GTGCGCAAAC ACTTAGATGA CTATTTTAAT    9480

AGAAGAAATA TTCGTCCAAA TGTCGTTGTA GAAACAGATC GATTCGAATC AGCAGTTGGA    9540

TTTGTTCATC TCGGCTTAGG TTACGCTATC ATTCCGAGAT TTTATTACCA ATCATTTCAC    9600

ACGTCTAATT TAGAATATAA AAAAATTCGT CCAAACTTAG GCCGAAAAAT TTATATCAAT    9660

TACCATAAAA AACGCAAACA CTCCGAACAA GTACATACAT TCGTACAACA ATGCCAAGAT    9720

TATTTATATG GACTTTTAGA GGCTCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG    9780

CTCAGTCAAC TGTATACCTT TTGCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG    9840

CTCAGTCAAC TGTATACCTT TTGCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG    9900

CTCAGTCAAC TGTATACCTT TTTCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG    9960

CTCAGTCAAC TGTATACCTT TTGCCTTTAA CTTAAGTTAT TAGTGCCTCT TATGTAGTTG   10020

CGTAGTCAaC TGTaTACCTT TTGCCTTTAA CTTAAGTTAT TAGAGCCTCT TATGCAGTTG   10080

CGCAGATCAT CGTATAAAAA TTAATGACGT CATTTCAAAA ATCGATACAA AAATAATTTA   10140

TTATAAAAAT TCTAAGAAAG AAGTGAAGCA GATGTTAAAA TCTATTAATC ATATATGCTT   10200

TTCAGTCAGA AATTTAAACG ATTCAATACA TTTTTATAGA GATATTTTAC TTGGGAAATT   10260

GCTATTGACT GGTAAAAAAA CTGCTTATTT TGAGCTTGCA GGCCTATGGA TTGCTTTAAA   10320

TGAAGAAAAA GATATACCAC GTAATGAAAT TCACTTTTCA TATACACATA TAGCTTTCAC   10380

TATAGATGAC AGCGAATTTA AATATTGGCA TCAGAGGTTA AAAGATAATA ACGTGAATAT   10440

TTTAGAAGGA AGAGTTAGAG ATATTAGAGA TAGACAATCA ATTTACTTTA CCGACCCTGA   10500

TGGTCATAAG CTAGAATTAC ATACTGGCAC ACTTGAGAAC AGATTAAATT ATTATAAAGA   10560

GGCTAAACCA CATATGACAT TTTACAAATA AGGTGTCATT ATAAAAAGGC CTCTTGAACT   10620

CCGTTAAAAT TTTAATTAAT TATTATATAA TAAGAGAACT TTTCAAACAA TACAGTTGTT   10680
```

```
TTACTGCAAT TATTTTTCAA ATATATCAAC GTTAATATAA CTTCTATTAA GAAATACTCA    10800

CATTCTGCCC TGCAATGCAA ATCTCGTCAC ATATAAATAT TTTTAATTAT TTTAAAAAAT    10860

GATGCACTAA ATTAGCAACG AGCTTAGCAG TTCTATTGTC AGCGTCATAT GTTGGATTCA    10920

TCTCAGCAAT ACTAACTGAA GACACCTTAT CACTTGGAAT AATACGTTTT GCTAATTCAA    10980

GAACAGTATG TGGATACAAA CCTAACACTG CCGGCGCACT TACCCCAGGC GCAAACGCAC    11040

TATCAATGAC ATCCATACAA ATCGTAAACA TAATGACATC ATGTTCATGT ACAAAACGTT    11100

CAATCATATC TTTAATTGTT GGTGATACGT GACTCAATAA TTCATCTGCA AAGACATAAT    11160

CAATCTTTTT CTCTTTAGCA TAATCAAATA AACTTTGCGT ATTACCACCT TGAGCAATAC    11220

CAAGCACTAA ATAATCTGTG TTTTCATCTT CTTCTAAAAT TTGTCTAAAG CTCGTTCCAG    11280

ATGTAGATTG TTGTTCAGCA CGTGTATCAA AATGCGCATC AATATTTATC ACACCAATAG    11340

ATTGTGTTGG ATAGACTTTA CGTGTTGCTA AATATTGAGC ATACGCAATA TCATGTCCAC    11400

CACCTAATAA AAATGTTTGT CTATGATTAG CAATTGACTT CGCTGCAAGC ATAGCAAATT    11460

CTTTTTGAGT ATCAATTAAT TCCTCATGAT CATGATAAAC ATTTCCGTAA TCGACTAAAG    11520

TTCACATTGA TTCAAATCCG GCAAACCTGC AAATGCTTGT TTAATCGCAT CTGGTCCTTC    11580

TTTTGCACCA ATGCGCCCCT TGTTTAAAGC AACACCTTTG TCAACAGCAT AGCCTAATAT    11640

ACCGACCCCT GATGGCATAC TACTCTTTTC CAGCTTAGAC AAATCTTCAA ATGTTACTGT    11700

TTGAAAATGT CTAAATTTTT TCGGGTCTGT TTCACTATCT AACCTTCCAG TCCATAAATT    11760

TGGTTCACCT TGCTTGTACA CAGCATTTCC CCCTCTTATT TATGTGGCTT ATTAACAATT    11820

AAAGTATAAC GTATAGGAAA TTTTGAATTC AATTCATAGT TAAATCCGTA TCTTAAAAAT    11880

ACTTATCTAC ATTACTTTTA CCCCTATTTT CTATGTAATA ACGAATACTT AGCTGATTTA    11940

TGTTAATAAA ATACGTCAAG ACTATTACAT TTTCATTAAT ATTGACATAG ACAATTTATC    12000

TCTCGGCTTG TAATATGTAT AATTGTTACT AAAAGATATT TTGCTTGTTA CCTAATGGAG    12060

GTTACATATA ATGAAGAACA ATAAAATTTC TGGTTTTCAA TGGGCAATGA CGATTTTCGT    12120

CTTCTTTGTC ATTACAATGG CGTTATCCAT TATGCTCAGA GATTCCAGT CTATAATTGG     12180

TGTCAAACAC TTTATATTTG AAGTTACAGA TCTAGCACCA TTAATTGCTG CAATCATTTG    12240

TATACTCGTT TTCAAATATA AAAAGGTCCA ACTTGCAGGT TTAAAATTCT CAATCAGCCT    12300

GAAAGTAATT GAACGTCTAT TGCTAGCTTT AATTTTACCT TTAATTATTC TAATTATTGG    12360

TATGTACAGC TTTAATACAT TTGCAGATAG CTTTATTTTA TTACAATCAA CAGGCTTATC    12420

AGTACCTATT ACACACATTC TGATTGGACA TATTCTGATG GCGTTCGTAG TAGAATTCGG    12480
```

232

```
TGTTGTTGGT TTGATGTATT CAGTTTTCTC AGCAAATACA ACTTATGGTA CAGAATTTGC    12600

TGCTTATAAC TTCCTTTATA CATTCTCATT CTCTATGATT CTTGGTGAAT TAATTAGAGC    12660

GACTAAAGGA CGTACAATTT ATATTGCAAC GACATTCCAT GCTTCAATGA CATTCGGACT    12720

TATTTTCTTG TTTAGCGAAG AAATCGGCGA TCTATTTTCA ATCAAAGTCA TCGCCATTTC    12780

AACAGCAATC GTTGCAGTAG GATACATTGG TTTAAGCTTA ATTATCCGAG GTATTGCATA    12840

TTTAACAACA AGACGAAACC TTGAAGAACT TGAGCCTAAT AATTATTTAG ACCATGTCAA    12900

TGACGATGAA GAAACTAATC ATACTGAGGC TGAAAAATCT TCTTCAAATA TTAAAGATGC    12960

TGAAAAAACA GGTGTAGCTA CTGCATCAAC GGTTGGTGTT GCTAAAAATG ATACTGAAAA    13020

TACAGTGGCT GACGAACCAA GCATTCATGA AGGTACTGAA AAAACAGAAC CTCAACATCA    13080

CATAGGTAAT CAAACTGAAT CTAATCATGA TGAAGATCAt GACATCACTT CGGAGTCAGT    13140

AGAATCAGCm GaATCAGTTA AACAAGCACC ACmAAGTGAC gATTTaACAA ACGATTCAAA    13200

TGAAGATGAA ATAGAGCAAT CATTAnAAGA ACCTGCGACT TATAAAGAAG ACAGACGTnC    13260

ATCAGTTGTA ATTGATGCAG AAAAACATAT CGAAAAAGCT GAAGAnCAAT CTTCAGATAA    13320

A                                                                   13321
```

(2) INFORMATION FOR SEQ ID NO: 5:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 8549 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
ATGTGTTGTA AACTTTTATG TTGAAAAAGC TACTTATCTC AATGAAAACA AGTAGCATTT      60

AATAAATTAA TTAGTATACA GCTAGTTTTT CTAATTGTTC TTTAACTTGA ATTAAGTTTG     120

ACCGTATTAG AGAGGCAGAT TGATCCATCG TTTGAATTGC TTGTCCTTCA TTTTCGTTCA     180

AGCCATTACA AACAACTTCA AACTGTTGTG CCATTTGATC AAGACGCGCA TGAGCTTGTG     240

TGTTTAAAAT AAACATATCG TCATAATGTG ATGGCGAATA GATAATTCGT CGTTGTATAC     300

AAACGTATAA AAACCTTGTC ATATCAACGG TTTTGGCATT TTTAAACCTC TGTGTTTTCC     360

ACGCATGTTT GCCCTTATTT AAATAATTTG CCCTTTTTTC GCCCCGAAAA AAAAACACAA     420

AAAAATAACC ACACTCCTAA ATTAATAGGT GGTGTGGTTT TGTTGATTGT AGGGGTATAA     480

AAATAACCGC ATTATTAAAG ATACGGTTAC TCTGTTATCT GTAAATATAA TAGTAGTTTA     540
```

```
AAACAGGACT  CCACATAAAA  ATCAACTCCT  TTATATACCA  TAATGATACT  ATATTTTCTA      660

GTTTATTTCA  ATTTTTCAGT  TTTTAAAAAT  GAGTTTCTGT  TTTTATTTAT  ACGCTTTTCT      720

GTTTTCTTTT  TAAATTTTAT  CTTTTTGTTA  TTCCATTCAT  TGTAAAATTC  TATTAAATTA      780

ACATAAAATT  TTTCATGCCC  TATTTTATTT  GTTGATGAGA  TATCAATGTA  AAGACTCAAT      840

ATTGTTTTTA  AATAGATTTG  ATGCAACGAC  TGATAAACCG  TATTACTATC  TGCTATGTTA      900

TTGGTAAAAT  GCATAGAAAA  ATATTCTAAT  TTATTCATGC  AATATATATG  GGTTTCATTA      960

TACTTCTTAA  TGAGTGTATT  TATACCTTGC  AATACGTCAT  TACTTTTAAT  AACAATTTCT     1020

TTTTCACCTG  TCGAAAAGT   CCACTGTTTA  TCTCCTATAT  TTTCTTTAAT  TGTTTTCTTG     1080

TTGTCAAATT  CTAAAATTAT  AGCCCGTAAA  CACTCTTCTT  TATAATTCTC  GTTCTTGAAA     1140

GTACGAAGCA  AAATTTTTAT  AAATTCGGTA  TTGGTGACTT  TTTTATAAGT  GTGATATTTT     1200

GCAATCTCTT  TATCAGTAAA  GACTGTTCTT  AGTTCGTGAT  TATCAAAACT  TAAATTCATC     1260

TTATTCTCTA  ATTCATTAAT  TTTATCTTGC  AAACCAACAT  TTTCTAAAAT  TTTCTTGTTT     1320

ATCTCCCCTA  TATCAAAACT  CCTTTTCGAA  ATTAATTTTG  AAAACTCGTC  TGCCATTTCA     1380

ACAGCCTTTT  CTTTCCTTTT  ATACCTTTTG  TTAAATTTAT  GAACCACCGT  TGCAGCATAA     1440

TACGATATCC  CACCAGATAA  AATAGATGaT  ATTATCGGTA  TGTATATATC  ACCTTTCATA     1500

TTTCCACCTC  TTTTAACACA  ATTAAGTATT  ATGATACACA  ACTTGCGCAA  AAAGATGTAG     1560

ACAGAACATA  ATGGCGAACA  AAAACAACCA  CCCAGTAACT  AGTATGGGTG  GCGTAgACTA     1620

TAACAACTCT  ATGTTATCAA  GATATATGTA  TCGAGTGATG  GCAAGGAAGA  AGTCTCCTGC     1680

GGGACCAACA  GTCAGATATA  TGGCCTCTGC  CGGGCTATAT  AGTTCACTCC  TACTATATAA     1740

AAGTAAGTAT  AACATAAAAA  GCACCCCGTA  AACTGTTATA  CGGGAATGCT  AAAGTCATAT     1800

ATACTACGGG  GAGTAGTATG  AAAACTATGC  TCTCTATCGT  AAGAAAAAAC  ACCCAGTGAC     1860

ATGCTTGGGT  GAACAAGGAT  AGATGTAAAT  AGTTGATGCA  TGTGTAcACA  TCATAACAAA     1920

AAACTAGCCC  GAAGcTAGCT  ATAACATAAA  AAAATAGGCA  AGTACCGAAG  TACCTGCCAG     1980

TTACGCACAT  TTAAATCTTG  AGAGTAATGT  TAAAAGTGT   ATAGGAATAT  TAACATCCAT     2040

CCAAATAGTT  ATTTAATAAC  TGTAAGATTC  CCTATAATTA  ATGTAGCaAA  ATTTTTATTC     2100

TAAGTAAATA  CTAAATCGTG  CTAAACTTAC  CAAAACTACT  TATTCTATTA  CCTGCCTTGT     2160

CTACCTCTCC  TGTCGCTATA  TAACGACGTT  GTCCACTATT  AGCAATATAA  GTAATCCATC     2220

TATAGCCATT  GATGCAATAT  GCGCCGTCAT  ATTTAATTGT  TGCGTTATTA  GGTAATACAC     2280

CTGTAATTCT  TGAATTAGTT  GAATAGCCGT  CCCTTACGTT  ATTACCTTTA  ACATTGGCAA     2340
```

```
CTGGCACTGG  TGGATTTTTT  TGGTTTTTAG  CTGATGTTTT  AACATTACCA  GCTACCAAAC    2460

CACCTATAGG  CTTACCATGA  ATCGCACCGG  CTATTAATTT  AGAATACAAG  TCATAGTTTT    2520

TCTTAATCCA  ATCCATATCA  TTTTTATTAG  TAATAAAACC  TAATTCAGAT  AAACGATAGT    2580

TTATATTTAT  TTCTGCTGAT  ACATTAACGT  TTAGTAAATC  ATTACGAGGT  GTTACACCTC    2640

TTATTTGTCC  TAAGTTATTT  TTAATAACAT  CTTGTATACT  TTTATCAATA  GTATCTGCAT    2700

TGAATTGACT  TGAAATAATA  ACATGCCCAC  CACTTGCACT  TTCTCCTGCT  GCGTCTAAAT    2760

GAATCTCTAG  AACAATGTCA  TACCCATGTG  ATTTAACCCA  ATATAAGCCA  TAATCTTTAT    2820

TATTTCCTAC  ATTAACACCG  TAAGCAGTAT  CTTGATACAT  ATCTTGTGAT  TGACTTGAGC    2880

CACCATATAA  TGCAACTTCG  TGACCTGCAT  GTCTTAAATA  CTTAGCGATA  TTTGGTGTTA    2940

TATATTTACG  GATAAAATCA  CGTTCATTTG  TTCCGTTTCC  GACTGCTCCA  GGATCGTTAT    3000

AACCATGACC  GGCTACAAGC  ATAATTTTTT  TAGGTTTAAT  TACTGCTTGC  TTTTTGGCAG    3060

TTGCTTGCTT  AATAACGCTT  TTAGCTTTAT  CTCCAACACT  TACTTTATCT  GGGAAATTTA    3120

ATCTAATAAA  ATACATTGGG  TCATCGTAAT  AATGAACATG  TCTTGTAACG  GTTTCGGGAC    3180

CCCAACCAGG  TTGCGCAACG  CCATTTGTCC  AACCTTTACC  ATTCCAATTT  TGGCCAAACG    3240

ATGTGAAAGT  GTTTAGATTA  GCGCTCTCAA  CAATTTCAAC  ATGTCCaGct  CCGCCACCAT    3300

ACTTTGACGG  GAAAACGACA  ATGTCCAACT  TTTGCGGTAA  AAAGCTATCA  TAGTTTTTAA    3360

TTATTTGCCC  GTATTTTTCA  ATCCTTGCTT  TATTATCAAA  TGGAATATTA  TAAGCGTATA    3420

AACCTTGTAA  CcTTTCGCCT  GTTGCTATCA  TAAAAAACAT  ATTTGCGTAA  TCGTAACACT    3480

GAAATCCATA  AAACAAATCA  GGATTGAACT  GCTTCCCTAA  TGAATTATCA  AACCATTTTT    3540

CTGCTTGGTT  TTTTGTTATC  AACATTGGTC  AACACCTACC  CTAAATCATT  TGTGTCGTTC    3600

ATATTCGTAG  GTGTCATTAC  TTCTTTAATT  GGCGCTTGCC  CTGTTGCTTT  TCTATACTTG    3660

TTTTCAGCTT  TATATTTCTT  TAGCTTTTGA  TTTGCCCATT  TACCTTCTTG  AGATGTTGGA    3720

TTATCTTTAT  ATGTAGTATA  TAAAGCAACA  ACTGTTAAGA  TAATCGATGA  AACACTTTCT    3780

TCATCTACTG  GTATCGGACT  TATACCTTTA  TTCGCTAAAA  ACTGATTGAC  TAATGCTAAG    3840

ATCAATACGA  TGTATCTTGT  TATTACTTTT  GCATCCATTT  GTTTGCTCCT  TTTATCCAAA    3900

ATAAAAAGCC  AGTGCCGAAG  CACTGACTCT  TAACTATTAC  TTACACTTAC  TAAACCAGAA    3960

ACACGACCAA  AAGCTATATC  CTAAAATTCC  CTTAAGCATG  GTAATCACCT  CCTTTAAATG    4020

CCAAAAATAG  TTTTTAACAA  GGCTATAACA  AATGTACTTA  GAATCGTCCC  TATTAATCCT    4080

AGAATCCACA  TCTTGATGTC  TCTAATATTT  TTAGCATTTT  TCTCTTTATT  TTTTTCATCT    4140
```

```
TGCGTTCTCA GACTGTCTTC TATTCTGTCG AATTTTTCAA ACATAGTCTT ATCATTTTCT    4260

TCTAATCGCG TTAAACGCCA ATCTTGTTCG TGTCGTTTGG TAAATCCAAA CATTACACCA    4320

CCCACTTTAT TCAAATTAAA AAGCCATAAG ATTATAACCT ATGACTCTAG ATTTTCTGGA    4380

TACTTTTCTC CTGTAATAAT TGCATATTCC TCTTTATCTA TAACTTCCAT ATCTACATAC    4440

CACGCTATAT CTTCTTTACT ATATTCTTTC AATTGATACC ATGTTTTAAT ATCTTCGAAT    4500

GTTGGTGAAA TTAATTTAAG CATTTTCAGT CTCTCCTTTA ACCTCTTCTA ATTTTTTATT    4560

AAGTGTCACA AGTTGTTTTG CCATTAGTGC ATTTTGCTTA TTAACTTGCA TCGATAACTT    4620

TGTACTTTGA ACAACTTGTT TCTGCATACT AGCAACCATT TTTCGTAAGA TGTCATCAGA    4680

AGCGACTGTG TTTTGTTCTT CACTGTCAAT CTGTTGATGC AAGTCATCTT TTTCTTCTGA    4740

ATAATCTTCG TTAAAAACTA TTTCCCCATT TGAATATTTA AAGGCTTTAG GTCTAAAAAC    4800

TTGAGAGAAA TTTTCTGGTA AATTTTCAAT ATCAATACCT TCTTCAAAGC CACCAATGAT    4860

AGCGTATGAA ATTATCTCAT TACGCTTGTT AACTAATATT TGCATTATTT TCTCACTCCT    4920

ATAATTTTGT TAATTGTCCC TCTATTTGCG TTCGCACCAG AGCCTCTTTG ACTTCCTAAG    4980

TCGAAATAGA CATCGTTTGA TATAGTTAAA GATGTACGAC TAGATTTAGT TAATCCAAAC    5040

TCATAAACAC CTCCACCATT TCCATCACCA TCTGGAAGAT TTGAGGGATT CAATGAAATC    5100

TTTCCTCCTC CAAAAGGACT GCCAAACTCT GTAAAGTCAC CACCTGGAAA AGTCCCATAA    5160

AAAATTAATA AAATAAATTG GTCTAAACTC TCATTTAAGT ACAATGTAGA GCCCACACCA    5220

TTTGCTGTTC CATCAAAAAT AACCGAATAC CTTTTATTAA ACTTGTCATC TGCGTATAAT    5280

TTAGCGTTAC TTTCGGCCAT ATTAGCTTTT GATTGGGCAC TTTGAACAGT TTCAAAAGGT    5340

GTATTGTAAT CATTAATAGC TAATTCTGAC CACTCAGACC ATGAACCCGC TTCTTTTCTT    5400

TTAACAAATA CTTTATTTGT ACCGTTCGGT CGATAAGTCA TACGCTTGTA ATCTGAAGTT    5460

ACTACTAAAT ATTCGACAGT ACCGTTAGTA CTAACACCTC TTGGATAATT TATAGCTTGC    5520

GAAACATAAA TAAATTGGGT TGAATCACCT ATTCTTTGTT CTGGATTATT AAAATCAAAT    5580

CCAGTAATCT GCATTATCTT ACCATCATCT TTAGTAATCT TAGCTTTTTG CCAATTTGAA    5640

GTAGAACCAC TTGTGACTAA ACCACCACTA TTCACTGACT GCTTGAAGGC TTCATGTTTC    5700

TCATCCATAT ATCGCTTTTG CTCATCGAAT GTTCTTGAAT ATGCTTGCGC TTTATTTTCC    5760

AAATCAGATA TATGGCTATT AGCAAGTTGC TTTAATTCAT CTATACTTGA AGATTTTGCT    5820

ATTTGAATAT CTGATAGACC TTTTTCTTTA GCTTTTTCAA TCAGACTCGC ATAATCTTCA    5880

CCATTTTTTA TAGCCTCGTC CATTGCTTTC GCACGATCCA TAATAGTTTT TTCTAATTCC    5940
```

236

```
TCAACGTTAA ATGTGATAGT TCTCTCGACA ACTACCACGT CTGAATTACC TAATTCTGCA    6060

ACCGAAACTT GAGCTTGATA ACTTCCATCT CGTTTAATTA CATCATTAGG TAATTGAAAT    6120

TTTAAAATAC CTTTAAATGG ATCTAATATT TCTAGTGGAG CAACTACCAT GACTCCTTTA    6180

CCTCGAATCG CTATTCGTGC kTTGATATTT tCTTCACTCA ATAATAACGG TTGATTATTT    6240

TTAGTGATAT TAAAAGAAG AACAGAAGAA TCACTCTCTC CTGTTCTAAA AGTTATATCT    6300

AGATTTGAAA TATTTCCATA ATGCGCTGTG TTTTCTAAAT TTATAGCTAC AGATTTCTCT    6360

AAATTACTCA TTAACTTATA ATTCTCCCTT CGTGTAAAGT CCATGGCCCT GAACTTGTTT    6420

TACTATCATA ATTTTTCAAT AGTATCTCAG CAGATGCTGT AACACTATTA CGAACTAGCC    6480

TATGAACAAA GCCACCTGTG TTTGAAGCTT CTACATATAA GTTCCAACCA GCTACCCCTT    6540

TACGTTCAGT TGGAAAATCT GTAAAACGTT TTGTATCATC CGTAGTTAAA TAAAACGACA    6600

TGCCTACTAT GTTAATATCT GACATTTTTG TGATGAATGA AGGTACTCTC TCCCATTTAC    6660

CACTATTTTT AGGCACATAA TTCCAGTCCG AAATGTCTCC AGTTCTTCCA GAAAGCACCC    6720

TTTCAAAAGT CATCATATTC CTTGCATAAC TATTACGCGT CAATATCTGA ATTACATCAC    6780

CGCCAGTTTG TGGTGGCTTA ACTTCCAAGA ACCAACCTGC ATCACGCCAT TCTCTTGGTA    6840

ATGGGAAATC ATCGATTTGA ACTGTATGAT CAGTGTATAA ATAGTAAAGA CCTGGCTCTG    6900

TTAACATCCC AAGATTCTTA AGTTTATCAG GCCTCATTGG TAAAGGTTTA ACTCTACCAC    6960

CTGTGTCACT CaTGATAAAA GGAACGCCTC TTGAGTGAAG TATTTCTAAA ATACCTCTTT    7020

GCCCAATCAT GAAAATACGA TGTGTTCTAT TTCCaTCACC ACCGACAGTA ACACCTAGCA    7080

TCAAAGCTTT TTTACCACTA TCTTTGTCAT AGTATATTTG CAAACCTTtC TgCTTCCGCA    7140

AATTCGCCAG GAAATGAATC tAgTGTTCCA CCATAGTCAG CATTAACCTG ATACGCTTCT    7200

TCTCCTGTTT CTAAATCGAA AGCCGTTAAA TAGTTTCTAT TATTTGGATT ACTGTCTCCT    7260

GTATACCAAT ACAAGTATTT TTCATCAAAA GTCACACCCT GCATTGGTTG GGTTTCGTTT    7320

GTTAGTCTCA TAGGGATACT GATTTTATGC AAAACTTTAT CAATATTTTT ATCAACATCG    7380

TCTAAACTTC TTATCTCTAT ATAAnTCATT GAGTTTTCAA GTTCCCACTG ACTTCTAGGT    7440

CTCTCaATTC TGTATAGAAT TTTATTTTCT TTTTCATTTA TGACAGGGGT GATGTAGGGT    7500

TTTTCTGGGT GTCCTGTAAA TACATCTTGC ATACCATACT TGCCATAGCT AATTTCCACA    7560

TTAGGCGTAT ACTTGAAACG AACTAATGTA TTCTCATTAT TACCATTTAA GATAAAACTA    7620

TAAATCCATA ACTCATcATC AATATATCTA TAACCGTTAT GTGTACCATG ACCCCCACCT    7680

ACAATCAATG AGCTGTCTAT AAATTGACCA TTAGGTCTTA GACGACTTAG CATATAGCCA    7740
```

```
ATTACTGCAT TTGTAAgAGG TGCAAGTTCT GTCACAAATA AAAATTCTTG CTTATCAGGT        7860

TCAAAACGAT ACTCGATATC AAGAATTTCT TGTTTGGTCT TATTTAATTC TCTTATAGTT        7920

TCCTCTTTAT TAATTTGAGT TTTGGTTTCC CAATCGTCTA AATGTTCTTT TAATGTGTCA        7980

AAGGTTTCGC CGTTTACATT AACTCGAGCT TGAACAATCT CATTAGCACT GTTATTACGT        8040

GGTGCCACAA CAAGTGCGTT AATTTGACTT TGTAAAGATT TGTTTACTGC TGCTTGCGAT        8100

CTACCATTAT AATAAATTTG CTCAGCGAAG TGTTGAATTG TTTTAGCTyT CTGATGCAAC        8160

TTAAACTCTG TTGTCAAGCC AAGCGCAAAT TGCTCTATTC TTTGTAAGTT TTGTATTTCC        8220

TTAGCTCTAT AATCTCGACC TGCTAAAGCT CCCAAATCCT TTATTAAATA CAAATTTTCC        8280

ATAATGCACC TTCCTTTCTA ATAAAATAGC ACTGTACCAA GTTTCCCACT ATCGTCAACT        8340

GTTATTTTCC ACAATTTACC GTTTGGGGAT TTCTGTACAA TGCTATTTTG AATAATTgcC        8400

TGctTCGCCT ATTTTTAAAT TATCTAATTT ATTTkTATCA TTTACCGAAA TGATACCGTC        8460

TTGAGGCAAT CCATCAATAn CACTACTGCC TGCATAAGGT ATCCCATTTA TAGCTTTCCA        8520

ATGTGTAGCT GGAAAGTACT GTTTATCGT                                          8549
```

(2) INFORMATION FOR SEQ ID NO: 6:

   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 3601 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: double
      (D) TOPOLOGY: linear

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
AGGCGTGTAG TGACTTACGG nTAGGAAACT ATGTATCCGA ATGATTTATT GAGACCAAAA          60

AGGCATTAAA GTCCATTGAA ATATCnGGTA GCGmGTTGGT ACgTGGACGT GGGGGCCCTA         120

GATGTATGAG TCAACCATTA TTCAGAGAGG ACATTTAACG TAATAAATTA TAGAmACGAG         180

GGTGAAAATA ATGACAGAAA TTCAAAAACC GTATGATTTA AAAGGCAGAT CATTATTAAA         240

AGAAAGTGAT TTTACCAAAG CAGAATTCGA AGGACTTATT GATTTTGCAA TTACATTAAA         300

AGAGTATAAG AAAAACGGTA TTAAGCATCA CTACTTATCT GGAAAAAATA TTGCACTACT         360

ATTCGAAAAG AATTCGACGA GAACGCGTGC TGCGTTTACA GTTGCGTCTA TTGATTTAGG         420

TGCGCATCCA GAATTTTTAG GAAAAAATGA TATTCAATTA GGCAAAAAAG AATCTGTAGA         480

GGATACTGCG AAAGTATTAG GTAGAATGTT CGATGGTATT GAATTCCGTG GTTTTTCACA         540

ACAAGCTGTT GAAGATTTAG CGAAGTTCTC TGGTGTACCG GTGTGGAATG GATTAACAGA         600
```

```
TCTAGAAGGA ATAAACTTAA CTTACGTTGG AGATGGACGT AATAATATTG CGCATTCATT      720

AATGGTAGCA GGTGCTATGT TAGGTGTTAA TGTAAGAATT TGTACACCTA AATCATTAAA      780

TCCAAAAGAG GCATATGTTG ATATTGcAAA rGAAAAaGCG AGTCAaTATG GTGGTyCAGT      840

CATGATTACG GATAATATTG CAGArcCAGT TGAAAaTwCm GATGCTATAT ATmCAGATGT      900

TTGGGTATCG ATGGGTGAAG AAAGTGAATT TGAACAcGTA TTAATTTATT AAAAGACTAT      960

CAAGTGAATC AACAGATGTT TGATTTAACA GGTAAAGATT CAACGATATT CTTACATTGT     1020

TTACCAGCAT TCCATGATAC AAATACACTT TATGGACAAG AAATTTATGA AAAATATGGA     1080

TTAGCTGAAA TGGAAGTTAC AGACCAAATC TTTAGAAGTG AACATTCAAA AGTGTTTGAT     1140

CAAGCTGAAA ATAGAATGCA TACAATTAAG GCAGTAATGG CAGCAACATT GGGGAGTTAA     1200

TCACTAAATG GAACGATATG AATATGATGT GTCTGATGAT ATAAGTGTCA TGTACAGACA     1260

CCTCATATTG GTATTAAAGG AGAAATGAAT ATGAACGAAT CAGGAGATAA CAAACTCAGT     1320

AAATCTTCTT TAATTGGACT AGTTATAGGA TCCATGATTG GTGGCGGTGC GTTCAATATA     1380

ATGTCTGATA TGGGCGGTAA AGCCGGTGGA TTAGCCATTA TTATTGGTTG GATTATTACA     1440

GCTATAGGAA TGATTTCATT AGCGTTCGTA TTTCAAAATT TAACCAATGA ACGGCCGGAG     1500

CTAGACGGTG GTATTTATAG TTATGmTCAA GCAGGATTTG GCGATTTTGT AGGATTTATC     1560

AGTGmTTGGG GATATTGGTT CTCAGCGTTT TTAGGCAATG TTGCCTATGC AACACTATTG     1620

ATGTCAGCAG TAGGTAACTT TTTCCCGATT TTTAAAGGAG GCAACACATT ACCAAGTGTT     1680

ATTGTCGCCT CGTTACTACT CTGGGGTGTC CATTTCTTGA TTTTAAAAGG CGTTGAAACA     1740

GCAGCATTTA TCAATAGTAT TGTTACTGTT GCAAAGTTAA TACCGATTTT ACTTGTAATC     1800

ATATGCATGA TAATTGCATT CAATTTTGAC ACTTTTAAAA CAGGCTTTTT CAGTATGACG     1860

TCAGAGGGTG TATTGCCATT TAGTTGGGCG AGCACAATGA GCCaaGTtAA AAGTACGrTG     1920

CTAGTGACAG TTTGGGTGTT TATCGGTATC GAAGGTGCAG TAATTTTTTC TAGTAGAGCT     1980

nAAAATGAGA AAGATGTAGG TAGTGCCACG GTTATAGGAC TTATATCAGT TTTAATTATC     2040

TATyTCTTAT TAACTGTATT AGCTCAAGGC GTGATTTTGC AAAATCATAT TTCGCAATTA     2100

GATTCGCCAA GTATGGCACA GGTGCTTGCA ACTATTGTAG GTGGTTGGGG ATCTACACTT     2160

GTAAATATTG GTTTAATTAT TTCGGTACTA GGTGCATGGT TAGGATGGAC ACTGCTTGCT     2220

GGTGAATTAC CTTTCATTGT TGCAAAAGAT GGATTATTTC CAAAATGGTT TGCTAAAGAA     2280

AATAAAAATG GAGCACCTGT AAATGCACTG CTTATTACCA ATATATTAGT ACAATTATTT     2340

TTAATAAGTA TGCTATTTAC ACAGAGTGCG TATCAATTTG CATTTCACT AGCATCAAGT     2400
```

CGACAGCAAG CAACTACTAA ACAATGGACG ATTGGTATCA TAGCCTCAAT TTATGCTATA   2520

TGGCTTATAT ATGCAGCAGG TATCAATTAC TTATTATTGA CGATGTTACT TTATATTCCA   2580

GCTCTTCTTG TTTATACaAT CGkTCmAAAG rATwATCAGa CACGTTTGAT TAAATCAGrC   2640

TATATTCtTT TTATGATTAT tATCGTACTT GCAGTTATCG GGTTAATTAA GTTATTGATG   2700

GGAACGATAA ATGTTTTTTA AAAGGAGCGA CAAAAATATG AAAGAGAAAA TTGTCATTGC   2760

ATTAGGCGGT AATGCGATAC AGACAACAGA AGCAACAGCT GAAGCACAAC AAACAGCTAT   2820

TAGATGTGCG ATGCAAAACC TTAAACCTTT ATTTGATTCA CCAGCGCGTA TTGTCATTTC   2880

ACATGGTAAT GGTCCACAAA TTGGAAGTTT ATTAATCCAA CAAGCTAAAT CGAACAGTGA   2940

CACAACGCCG GCAATGCCAT TGGATACTTG TGGTGCAATG TCACAGGGTA TGATAGGCTA   3000

TTGGTTGGAA ACTGAAATCA ATCGCATTTT AACTGAAATG AATAGTGATA GAACTGTAGG   3060

CACAATCGTT ACACGTGTGG AAGTAGATAA AGATGATCCA CGATTTGATa ACCCAACTAA   3120

AcCAaTTGGT CCTTTTTATA CGAAAGAAGA AGTTGAAGAA TTACAAAAAG AACAGCCAGA   3180

CTCAGTCTTT aAAGAAGATG CAGGACGTGG TTATAGAAAA GTAGTTGcGT CACCACTACC   3240

TCaATCTATA CTAGAACACC AGTTAATTCG AACTTTAGCA GACGGTAAAA ATATTGTCAT   3300

TGCATGCGGT GGTGGCGGTA TTCCAGTTAT AAAAAAAGAA AATACCTATG AAGGTGTTGA   3360

AGCGGTTATA GATAAAGATT TTGCTAGTGA GAAATTAGCA ACGCTGATTG AAGCAGATAC   3420

CTTAATGATT CTTACGAATG TAGAAAATGT ATTTATTAAC TTTAATGAAC CTAATCAACA   3480

ACAAATCGAT GATATTGATG TAGCAACACT GAAAAAAtAC GCGGCACAAG GTAAGTTTGT   3540

GGAAGGATCG tGTTGCCAAA AATAGAAGCT GCGtACgtTT GTTGAaAGtG GGGaAACCAA   3600

A   3601

(2) INFORMATION FOR SEQ ID NO: 7:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 573 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGACACTATT AAATGAATTA GAGCACAATC TAACAAATCA AATTCATTTT TCAAAAGATG   60

AACGACTCAC ACATATCGCT TTAAAGTTAT TCGAAACAAC CGATCCTGTT TCAACAAAGC   120

AACTTGCGCA AGATGTTAAT GTTTCGCGTC GGACAATTGC AGATGATATT AAAATGATTC   180

TTATTGGTGA GGAAGATCAT TATCGTAAAG CGTATGCACA CTTTATACAT CAATATATGA   300

AACAAGCTGC ACCTTTTATA GAGGCGGATA TCTTTAATTC AGAATCAATC GCATTGGTTC   360

GCCGTGCCAT TATTAAGACA TTAAATAGTG AAAATTATCA TTTAGTTCAG TCGGCTATCG   420

ATGGCTTAAT CTATCATATA CTCATTGCCA TTCAGCGTTT AAATGAAAAT TTTTCGTTCG   480

ATATACCTAT CAATGAAATT GATAAATGGC GACATACTAA TCAGTATGCn ATTGCTTCAA   540

AAATGATAGA AAACTTAGAA CGCAGTGTAA TGT   573

(2) INFORMATION FOR SEQ ID NO: 8:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1221 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTGATATTTA TAACGTTATA TTTTAATAGT TCACCTGGAT TATTAAATAA ATAGTCCGCC   60

AAATTTTCTT TTTCTTTATC AATCTGaTkG TAATTAACaC TTTCGaCTTC TGTAGGAATT   120

CTAATGTCAA CAGAAGCATT GATATAAGCT TGATGTTGCA TGCAATCACA CTCCTAATCC   180

TTCATmTmAA ACGGAGAAGT AAACCCGTCA CTATTCAAAT TCAATCCTTT TGCCCAATCA   240

ACAGGCTTAT TCATGATAGT TTCGATTTCC TTAAGTCCAT TTGAACCTCT AGGTATTTCT   300

ACAATTACTT CATCATGGAC ATGGCCAACT ATTTTAAAAC CTAATGCTTC AAGCCTTGCT   360

ATAGAAATCG CAAGTAAATC CCTTGCAGTT GCTTGAACAA TATTCTCGAC TAACTTCCCA   420

CCATACGTTT TTAACTTTGA CCATTTACGG TTAAGATCTA ACCCCATAAA TTCAACAACT   480

TGACTACCCC AACTATTTTC ACCAACTAAA GCTTTTGGAT AAGCTAAAGC TCTTCCACTA   540

GGCAGTTCAA TCATTAGAAA ACCTTTTTTC ATATAAAATC TAAGTCCATG TGTATGATGC   600

GTCTTTCGGG ATTTTACAGT ATTAATTGCA GCCTCTTGGC AAGCCTTCCA AAAATTAACT   660

ATGTTAGGAT TTGCGTTACG CCAACTATCA ACTAAACCTT GTAACTCGTT TTCTTCAATG   720

CCCATTTCCA ATGCACCCAT TGCTTTTAAA GCTCCAGCGC CACCTTGATA GCCTAAAGCT   780

AATTCGGACA CTTTTCCTTT TTGTCTGAGA GGGTCGCCTT TAGTTATGCT TTCTACCGGT   840

ACATTAAACA TTTGAGAAGC CGATGCTTCA TATATCTTTC CGTGTGTGTT GAATACATCT   900

AAACGCCATT GTTCTTTTGC ATACCATGCT ATGACTCTTG CCTCTATTGC AGAAAAATCA   960

CTTACTGCTA GTTCATTACC TTCTTCAGCA GTAAATGTCG TCCTAACTAA TTGACTTAAT   1020

```
AGATCTCTTG CTATTTCTAA TTCAGTATCT GAAATATAAT GCTTTGTTAA ATTCTGAAGT      1140

TGTACACCTC TACCTGCCCA TCTTCCAGTA CCGGCACCGT AAAAATTGAAA CAGACCTCTT      1200

ACCCGTTCAT CACTGCACAT C                                                 1221
```

(2) INFORMATION FOR SEQ ID NO: 9:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1090 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
TTTTGTTTGG TATGAGGTAG CAATGACGAC GTGTCATTGG TGGAGATTGT AAAAATACAT        60

AATAAAAAGA AGCGGCAATG TATACCGCTC CTTTTTTATA CTACATACCG ATTTTCAACC       120

ATCTCTTTCT ACTTAGTAAT AAGACAATAG TATTAACTAT AAATAGAAGA ACGAAGAATG       180

ATACTATATT TATAATTTCA GTAGGACACA TAAATGTTGA CTCGTTATTC AATATTTTTT       240

CTACGGCACG ATACATCGTA TTGCTCGCCT CAAATGGAGC AACGATACCA AATATATTTT       300

TATTAATGGC AACTAAGATG ACTGAACCAA TCCAATATAC AATGCTGATA CCTAAGCTGA       360

TTAAAATGTT AGGTGAAACC ATACTAATCG TTCCAACAAC TAAGATATAT TGTAAGATAA       420

CGAGTGAAAA TAAGATTATT AATAGTAAGT AATGTGAGAA ATCCGAATAT ATAATTGAAA       480

TAATAGTGAT ACTTAGAATT ATGAACACTA AACATTCAAA AAATAACACT GCTACCTTTT       540

TATAGAAGAA GGTAAAGATA TTATCGCCAA TCAATTTATA AAACAGGATA TTTTTATTCG       600

AATACTCTTT ATTAATAAAA TATGCAATAA CAAATGAAAA TAGTAAGAAC CCTAATTGCG       660

TTGCAACAGT ATATGAACTG AAGAAAAACT GGCTATAGCT TAAACTTTTA ACTTTGTCTA       720

TACCTATTGG TAAAAAATAC CCAAGTAAGA AAAGGAATGT GAATAGCACA ACAAGCGTGT       780

AAATAATTTT ATTGGAAATA CTTTTTTTAA ATTCTAATTT CAAAGTGGAC ACCTCAATTA       840

TAAATTAATG TAATCATTTA TGACTTCTTC TTTTGATTGG TACTCTTCTA TTTGAAGGTC       900

TTTAAAAATA AAGTATTTAC CCGGCAAAGC ACTTAAATCG GATAAATTaT GTGTAATATT       960

GATAATAGTT TTAGTTTGAT GGCTTTGAAT AAAATCATTT AAAAATTCAT AAATTTCATT      1020

AACTGTTTTC TTGTCTAAAG CGTTTGTAAC TTCATCTAAT ATGATTAAAT CATGATCTTC      1080

CAATAAGAAA                                                             1090
```

(2) INFORMATION FOR SEQ ID NO: 10:

(A) LENGTH: 904 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TTAGGACTAT TTTATCATAT TCATTTAAAT TACGGCTAAA AATTTTAAAA ACGGGGATTA   60

ATATATGGAA TTAAGCTATG AAAGTTAATT GATACTTGCA TTTTACGCTG ATTTATATAA   120

GAATAACTAT TGTATAGTTT TAAAAACGAA CGTACGTTTG CAGGAGGCGA AATCATTGGC   180

AATGAATAAA CAAAATAATT ATTCAGATGA TTCAATACAG GTTTTAGAGG GGTTAGAAGC   240

AGTTCGTAAA AGACCTGGTA TGTATATTGG ATCAACTGAT AAACGGGGAT TACATCATCT   300

AGTATATGAA ATTGTCGATA ACTCCGTCGA TGAAGTATTG AATGGTTACG GTAACGAAAT   360

AGATGTAACA ATTAATAAAG ATGGTAGTAT TTCTATAGAA GATAATGGAC GTGGTATGCC   420

AACAGGTATA CATAAATCAG GTAAACCGAC AGTCGAAGTT ATCTTTACTG TTTTACATGC   480

AGGAGGTAAA TTTGGACAAG GCGGCTATAA AACTTCAGGT GGTCTTCACG GTGTTGGTGC   540

TTCAGTTGTA AATGCATTGA GTGAATGGCT TGAAGTTGAA ATCCATCGAG ATGGTAATAT   600

ATATCATCAA AGTTTTAAAA ACGGTGGTTC GCCATCTTCT GGTTTAGTGA AAAAAGGTAA   660

AACTAAGAAA ACAGGTACCA AAGTAACATT TAAACCTGAT GACACAATTT TTAAAGCATC   720

TACATCATTT AATTTTGATG TTTTAAGTGA ACGACTACAA GAGTCTGCGT TCTTATTGAA   780

AAATTTAAAA ATAACGCTTA ATGATTTACG CnwGGgTAAA GAGCGTCAAG AGCATTACCA   840

TTATGAAGAA GGGAtCaAAG rGTTgTTAGT atGTCCAaTG ArGGAAAAGA AGTTTTGCCT   900

GACG   904

(2) INFORMATION FOR SEQ ID NO: 11:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 11271 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GATTTCTAAA TCAAGATCTG TTTTACGATA ACCATTCAAA CCTTGACGTT CATCTTCTTC   60

AGGTTGATTT TGTTGCTGTG TGTCTTTGTT GTCAGAAGTC GCTACTGTTT TTTTATTATC   120

TGTTTCTTTA GTCATAACAA ACGCCTCCGT TATAAAACGC TATATTTAAT GATATGTGAT   180

243

```
TTAATAAGAC DATTCAGCAA GTTTTAAAGT ATTATTTGAC TATGTTGGAT TAGGCATCTA        300

GTCCTATAAT ATCACTGACA TTGTCAAAAT GATGATCTTT TAAGTAACGT GCGATGCCTT        360

TGTTCATTTT CTTAGTTAAA CCTGGGCCTT CAATAACAAG TGATGAATAA ATTTGAATAA        420

GTGACGCACC GTGACGCATC ATTTTGATTG CATCTTCAGT ACTGAATACG CCGCCTGTAC        480

CTATAATTAA AAATTCACCA TTTGTTTGCT GATAAgCATa CTTAATCAAT TTTAAATTAC        540

GTTCAAATAA TGGACGACCA CTCAAACCGC CTTCTTCGAC TTTATTAGCA GAAGTTAAAC        600

CATCTCGTTG TCGCGTTGTG TTTGCTAAGA TGATACCGTC AAATGTCTCA GTAATCGCTG        660

GTAATAGTGC TTTTAAGCCA TCGAAATCCA TATCAGACGT TAGTTTTAAA TAAATTGGCA        720

CTGTTACATC ATGTTGTTTT TTAAATGCTG TTAAAGCTTG GCATAACATT GAAAATTCAT        780

CTTTATCATG GAAGTTTTGA AGATTTTCAG TATTTGGAGA ACTGATGTTG ACTGTGAAAA        840

ATGAAACGTC GTGTTTAAAC GTATCAATAA CCTTTATATA ATCTTGATAA CGCGCTTCAT        900

AAGGTGTCAT TTTATTCACA CCAACATTGA TACCAACAGG TACTTGATAA GCATTTTTAC        960

GCAAATGACT TAGTGCTTTG TTCATACCAA TATTATTGAA GCCCATTCGA TTTATCAAGG       1020

CGTCATCTTC TAATAATCTA AACATGCGTG GTTGAGGGTT ACCCGGTTGA GGTTTAGGTG       1080

TGATACCACC TAATTCTAAA GCACCGAATC CAAGGTGTTC CAATGCTTTT GGTACTTCGC       1140

AAGATTTGTC GAAACCAGCT GCTAAgCCAA TTGGATTGTC GTACGTATTA CCTTGTATCG       1200

TTTGTGATAA CGTTGGATTC TTATAAGTAA ATAGTTTATC GACGACTGGG AATAAAACCG       1260

GaAACTTTTG TaACGTTTTT AATGCATCGA TAGTTAGTCC GTGTGCTTTT TCGGGTTCGA       1320

TTTTGAATAA GAAAGGTTTA ATTAATTTGT ACATGAGTAT GCTCCTATTT CATTATATTT       1380

GAGGCTTACT ATCCTCAACT TAATATATGT GAAATATATT CTTTTAATAG ACTAGCATTT       1440

CCATÂCATAA TTTCCTAGTT AAAACTAAAA AGTTTTGAAA ATTGACGCAA gTTTGAATAA       1500

CGTTTTTAAG ATTAAATCAT CCTAATTAGG CAATATTATA GTATAAAGTA AGTAGATTGG       1560

AAGGTGTTTG TATGAATGAA CAATGGTTAG AGCATTTACC TTTAAAAGAT ATTAAAGAGA       1620

TTTCACCAGT GAGTGGTGGT GATGTAAACG AAGCATATCG AGTCGAAACA GATACGGATA       1680

CATTTTTCTT ACTTGTCCAA CGTGGACGTA AAGAATCATT TTATGCTGCA GAAATTGCAG       1740

GTTTAAATGA ATTTGAACGT GCAGGTATCA CGGCACCTAG AGTAATTGCA AGTGGCGAGG       1800

TTAACGGTGA TGCGTATTTA GTGATGACGT ATTTAGAAGA AGGGGCTTCA GGGAGTCAAC       1860

GCCAATTAGG GCAACTCGTA GCTCAATTAC ACAGTCAGCA ACAAGAAGAA GGCAAATTTG       1920

GCTTCTCATT ACCTTATGAA GGTGGCGATA TTTCTTTTGA TAATCATTGG CAAGACGATT       1980
```

```
GGCTATGGGA TGCCAACGAT ATCAAAGTAT ATGACAAAGT GCGACGTCAA ATTGTGGCGG      2100

AATTAGAAAA GCATCAAAGT AAACCGTCTT TATTACATGG TGACCTATGG GGTGGTAATT      2160

ATATGTTCTT ACAAGATGGT CGTCCGGCGT TATTTGATCC AGCGCCATTA TATGGTGACA      2220

GAGAATTCGA TATCGGTATT ACAACGGTAT TTGGTGGTTT TACGAGCGAA TTTTATGATG      2280

CGTATAATAA ACATTATCCA CTCGCAAAAG GTGCATCCTA TAGACTTGAA TTTTATCGTT      2340

TATATTTATT GATGGTCCAT TTATTGAAAT TTGGTGAGAT GTACCGTGAT AGTGTTGCGC      2400

ATTCTATGGA TAAGATTTTA CAAGATACAA CAAGTTAGTT AAGACGTTAG ATTGAGATAA      2460

ATAGATAATA TGCACAGATA TTTTTACAAT GAGAAGCGAT ACAGCTGCCT CAATAAAAAT      2520

ATTTGTGCGT TTTTATTGTT GGAAAATAAA ATTTTAATCG CTATTGTTAA TTTCTGTAAT      2580

GTAAAACAAG GTTGAGTTAC AATAAAAGTG ATTTTATAAC TTTTTGTTCA ATAAAATTCT      2640

AGGAATGATA CATATTTATT GATACAATAA TTTTGAATAT AATCATAAAA CAATATTTAA      2700

GTATAATTGA ATGTTTGAAT ATCATATATT GATACAGTTT CTAATAATTT TAAAATAATT      2760

TAAATGGAGA GAGGTGTAAA TGATGAGTAC AGTTCAAAGT GATATTTTTA AGACCAATAG      2820

TGCATCATCA TCTATTAAAA GCGCTGTTGA AACATGTAAT AATGTGTCGA AACCGGATAA      2880

AGATGAAAGT ACAACAGTAA GTGGAAATAA TAATGCTCAT AGTGTGATAG ATGATTTGAT      2940

GAGTAAGAAT CAATCTGTTG CTGAAGCAAT ACGAACTGCG AGCGATAATA TACAAAAAGT      3000

TGGTGAGGCT TTTGACCAAA CTGACGTAAT GATTGGTAAT GAAATTGGTA AAAATTAAAA      3060

CGTGGTGAAA TGATGTCGAA TAAACTGGAT GAAATCAATA AAATAATCAC AGCGAAACAT      3120

GAGCAAATGG ATGACTTATA TGATGAAAAG CGAGAGGTTA AAGCATTGAT AGATGAAAGT      3180

GATGCGCTTA ATCATTCGAT AGATCAATTA TATCAACATT TAGGTGAGCG TTATTATAGT      3240

AGCAATATGG CTAGTCGTAT GGAACAGTTC CGCGATGAAT TTCATTTTGC GAAACGACGT      3300

TCAACGGAAG CGTTATACGA GCAGCAACAG CAAATTCAAC ATGGCATTCG TAAAGTGGAA      3360

GAAGAGATGA TTGACTTGGA AATGCGAAGG AATGTTGAAA TTGAGACGGT GACAAAGGAG      3420

GAAAATAAAT GGAAACAATA GGAAGCATTA TTTATTTAAA AGAAGGTTCG CAAAAGTTAA      3480

TGATTATTAA TAGAGGmCCA aTTGTAGAAA TTGAAAATCA AAAGTATATG TTTGACTATT      3540

CTGCATGTAA ATATCCGATT GGTGTTGTAG AAGATGAAAT TTATTATTTT AACGAGGAAA      3600

ATATAGATTC AGTTATTTTT AAAGGTTATT CTGATCAAGA TGAGGTTAGA TTTCAAGAGT      3660

TGTTTGAAAA TATGAAACAA AATTTGGATA GTGAAATACA ACGTGGAGAA GTTACACAAC      3720

AATAAAGAAA TACTTTTTCT TTATTGGGGT GGGACGACGA AATAAATTTT GTAAAAATAT      3780
```

245

```
ATGTCATTCA TAATCATTTG AACTAAACGT AGCAGCCTTA AATTTTAAAA AAAGACACAT      3900

ACCAACTTCC GAAATGTAGA TGAATTCTCT ACAATAACGG AAGTTTTTCT TTTAATATTG      3960

AAATTTCTCA AGGATAGGTC TATACTTTAT AAATCGTAAT TATTACGATT TATAATCAAA      4020

AACAATAACT TGAAATAGAT CATTGAGGGA GTGTTAATAT GCAACATCAT AAAGTGGCTA      4080

TTATcGGTGC CGGTGCTGCA GGTATAGGTA TGGCCATTAC CTTAAAAGAT TTCGGTATAA      4140

CAGATGTCAT TATTTTAGAA AAAGGAACAG TAGGACATTC ATTTAAACAT TGGCCGAAAT      4200

CGACCCGTAC GATCACGCCA TCATTTACGT CTAATGGATT TGGCATGCCT GATATGAATG      4260

CAATTTCCAT GGATACTTCA CCAGCATTTA CATTTAATGA AGAACATATT TCCGGAGAAA      4320

CATATGCTGA ATATTTACAA GTGGTTGCCA ACCATTACGA GCTGAATATC TTTGAAAATA      4380

CAGTTGTCAC AAATATATCT GTAGATGATG CATATTATAC GATTGCAACG ACAACAGAGA      4440

TATATCACGC GGATTATATC TTTGTCGCAA CAGGTGATTA TAATTTCCCT AAAAAgCCAT      4500

TTAAATATGG TATTCATTAT AGTGAAATTG AAGACTTTGA TAACTTTAAT AAGGGGCaAT      4560

ATGTGGTTAT CGGAGGTAAT GAAAGTGGCT TTGATGCTGC ATATCAACTT GCAAAAAATG      4620

GCTCTGACAT CGCACTTTAT ACTAGCACAA CCGGTTTAAA TGATCCGGAT GCTGATCCTA      4680

GTGTTAGATT GTCACCTTAT ACACGTCAGC GACTAGGTAA TGTCATTAAG CAAGGTGCTC      4740

GCATCGAAAT GAATGTACAT TATACAGTTA AAGATATTGA TTTTAACAAT GGACAGTATC      4800

ATATCAGTTT TGATAGCGGA CAAAGTGTGC TTACACCTCA TGAACCAATA CTAGCAACTG      4860

GCTTTGATGC AACAAAAAAT CCAATCGTTC AACAATTATT TGTGACAACA AATCAAGATA       4920

TTAAATTAAC AACACATGAT GAATCGACAC GTTATCCGAA TATTTTTATG ATTGGTGCAA      4980

CAGTTGAAAA TGATAATGCC AAATTATGCT ATATCTATAA ATTTAGAGCG CGATTTGCAG      5040

TACtTGCACA TCTTTTAACA CAGCGGGAAG GcTTACCAGC TAAACAAGAT GTCATTGAAA      5100

ATTATCAAAA AAATCAAATG TATTTAGATG ATTATTCATG TTGTGAAGTG TCATGCACAT      5160

GTTAGAAGTG AAATATGATA TGAGAACTGG GCATTATACG CCCATACCTA ATGAACCTCA      5220

TTATTTGGTT ATTAGTCATG CGGATAAACT TACCGCAACA GAAAAAGCGA AATTAAGATT      5280

ATTAATCATA AAACAGAAAT TAGATATTTC ATTGGCAGAA AGTGTAGTTT CTTcGCCTAT      5340

AGCGAGTGAA CATGTGATAG AACAATTGAC ACTATTTCAA CATGAGCGAC GACATTTAAG      5400

ACCTAAAATA AGTGCGACAT TTTTAGCCTG GTTGTTGATA TTTTTAATGT TTGCATTGCC      5460

AATCGGTATC GCTTATCAAT TTTCAGATTG GTTTCAAAAT CAGTATGTGT CAGCATGGAT      5520

AGAATATTTA ACTCAAACAA CATTGCTCAA TCACGATATA TTACAGCATA TATTATTTGG      5580
```

```
ATTGATTAGT TTATCAACTG CTATAATTGA TCAAACAGGA CTCAAATCAT GGATGATATG    5700

GGCAATTGAA CCGTCAATGT TATGGATAGG ATTACAAGGT AATGATATCG TGCCACTATT    5760

AGAAGGGTTT GGATGTAATG CAGCAGCTAT TTCACAAGCA GCACACCAAT GCCATACCTG    5820

CACGAAGACA CAGTGTATGA GTTTAATAAG CTTTGGTAGT TCTTGTAGTT ATCAAATAGG    5880

TGCGACATTA TCTATTTTTA GTGTAGCTGG AAAGTCATGG CTATTTATGC CGTACTTAAT    5940

ATTAGTACTT TTAGGTGGCA TCTTACATAA AGGATATGGT TGAAAAAGAA TGATCAACAA    6000

CTTAGCGTTC CGCTACCTTA TGATAGGCAA TTACATATGC CAAATATACG TCAAATGTTG    6060

CTACAAATGT GGCAAAATAT ACAAATGTTT ATCGTTCAAG CGCTACCTAT TTTTATCACA    6120

ATCTGTCTTA TTGTTAGTAT TTTATCACTA ACGCCAATTT TGAATGTTTT ATCACAAATA    6180

TTTACACCTA TATTATCGTT ATTAGGCATC TCGTCAGAAT TGTCACCAGG GATTTTATTT    6240

TCAATGATTC GAAAAGACGG CATGCTCTTG TTTAATTTGC ATCAGGGCGC CTTATTACAA    6300

GGAATGACAG CAACACAGTT ACTACTACTT GTGTTTTTTA GTTCAACATT TACAGCGTGC    6360

TCGGTCACAA TGACGATGCT TTTGAAACAT TTAGGTGGTC AGTCAGCACT AAAATTAATT    6420

GGAAAGCAAA TGGTGACATC ATTGTCTTTA GTTATTGGTG TAGGCATCAT TGTTAAAATA    6480

GTAATGCTGA TTATTTAAAA AAAATGAACT ATAACTGAAT ATAGAGTCAT GTCAGTCAAT    6540

AGGAGATCTA TCTTGGAATA TGCTATTCAT ATGAAGTATA AGAGGAGAGT CGCAGATGAA    6600

AATAGTTATT ATAGGTGGGT TTTTAGGTGG CGGTAAAACG ACTGTCTTAA ATCATTTGCT    6660

CGCTGAATCA TTAAAGGAAT CGCTGAAACC AGCAGTCATC ATGAATGAAT TTGGGAAAAT    6720

GAGTGTTGAT GGTGCCTTAG TATCTGAAGA CATACCTTTA AGTGAACTGA CAGAGGGGTG    6780

TATCTGTTGT GCAATGAAAG CAGATGTATC AGAACAGTTA CATCAATTAT ATTTAAAAGA    6840

GCAACCAGAC ATTGTATTTA TTGAATGTAG TGGGATTGCA GAACCGGTCT CTGTCTTAGA    6900

TGCTTGTTTA ACGCCTATTT TAGCTCCGTT TACAACAATT ACACATATGA TTGGTGTAAT    6960

AGACGCAAGC ATGTATAAAC ACATTAAATC ATTCCCTAAA GACATCCAAG GCTTATTTTA    7020

TGAGCAATTA GCATATTGTT CTGTCTTATT TGTTAATAAA ATAGATTCAG CAGATGTTGA    7080

AACAACGAGC AAACTATTGA AAGATTTAGA AGTTATTAAC CCAGAGGCCG ATATACAAGT    7140

CGGTATGCAT GGCAGCGTCA CTTTGCCAAT ATCAGTTAGA CAAATGACAG CAACTTCTGA    7200

CAATAAACAT AAGTCTTTAC ATCAAATGAT TAATCATCAA TTTGTGCAAT CACCAGTCAA    7260

ATGTACTAAA GCAGAGTTTA TAAAACGTTT AGCATGCCTT CCGTCTCATA TTTATAGGTT    7320

GAAAGGGTTT ATGACATTTG AAGACACCGC ACATACGTAT CTCATTCAAT TTACACAAGG    7380
```

247

```
CGGAAAGGGT ATTTCAAAAG AAGACTATCA ATGTTTGGAA CAGTAGTGTT TTCAGTGGAA    7500

GAGAATGGTT AACATGCCTT CATGTATAAT AACGAGTTGA TTTGAACGTT TAAGCGTAAA    7560

TAAAAATAAG CTTGGTCAGC CATCAAATAT AATTTGAAAA CTGTCCAAGC TGTTTTATTA    7620

GAGAACAATC AATTAACCCC ACATATTTAA TAATACATCA GCAAAGCCTT CAGGTTTTTG    7680

AATATAACCT AAGTGACCGC CTGGAATATC TACAATAGGT ATGCCAGTTT CTTTATTTAT    7740

ATAAAGTTA ACATCTTGTG GGAAGGAGCC TCTAGAATCT GTCCCATTTA GTAGGGTGAT     7800

TTTATCGCTG TATTTTGTGA AATCATCCAA AGTAATATCT GAATGCGTAT ATTGTCTAAT    7860

TTCAAATTCT GACCAGAACA TCGTACGTTT GTACTGTTCT ATACGTCCTT CTTCAGTATC    7920

AGCAGGTTGA GACATCATTT TTGCATCAAT TGGTGCGATA TTTAATGTTT CGCCAAATGT    7980

TTTCATGCCT TTTTCTAAGC CTTCTGTTAA AATTTGATGC ACAATGTCAT CATTTTTATC    8040

TTTCCAATAA GTACTGTCTG GTAAAAATGT ATTAATTGGT GGTTCGTGAA ATGCAATCTT    8100

TTTAACGACT TCAGGGTAAT CTTTTAACAC ATGCATCGCA ACGATTGAAC CTGAACTTGA    8160

ACCTAATATA TAGACAGGTT CATCACTTAA TGACTTTGCA AGTTCGGCAA TGTCCTGTGC     8220

GTCGCGTTTG ACACGATAAT CACTGTCAGG GTTTGAAGCG GAATCAGGGA GTGGTTCAGT    8280

TAACTCGCTT TCTCCATAAT CACGACGATC AACGGCTACA ACAGTAAAAT GGTCTTTTAA    8340

CTGTTCTGCA AGAGGCAGAA AAATGTCTCC GGTACCGTTT GCACCAGGAA TAAAGATGAG    8400

CACGGGTCCT TGTCCGACTT GGTGGTATCG TAATTTAGCG CCTTGTAATT CTAAAGTTTC    8460

CATATTCAAT GACCTCCATT TGTTAATTGT TAGGTGATAA ACCTAATAAT TTAGCACCAT    8520

TTGTATAACT TATTTTCTCT TTTTCTTCAT CTGTTAAACC CAGTTCATCT AAAAATACAC    8580

CTAATTTTTC AGGCTCAATA TATGGATAAT CAGCAGCATA AAGAATTCTA TCAATACCTA    8640

CTTCTTTCTT GACTAAATCA AACTGTGGCT TCGTTAACAT GCCACTCGGT GTGATATAAA    8700

AATTATTTTT AAAGTAATAG CTTACAGGGT GGTTCAAATG TTCAGCGAAT AAAGCTTCAT    8760

CCATACGTTC TAAGAAGAAT GGGATAAACT CACCCCAATG TCCAATAATC ATATTTAACT    8820

TTGGATAACG ATCAAAAATA CCAGATAATA CTAGATGTAT TGTATGAATG CCGACATCAA    8880

TGTGCCAACC ATAACCAAAA CAAGCAAATG TTGCCGCAGT TACTTCAGGA TAATTTCCTT    8940

TATAGTATGA TTGATAAATG TCACTGTTAA CTGGCGCGGG ATGTAGATAA ATCGGTACGT    9000

CTAAATTTTC AGCTGTTTTG AAAATAATGT CATATTTGTC TTGATCAAGA AAACCATCTT    9060

GTGCACGTCC CATAATGAGC GCACCTTTGA ATCCTAAATC ATTGATGCAA CGTTCGAATT    9120

CTCGCGCTGC GGCTTCAGGC TCATTGATAG GTAAAGTTGC AAAGCCTACA AAGCGATTGG    9180
```

```
TCTGACCAAC CAAATTTGAA GGAGAACCAT TTCCATAAGA TAAGACTTGA ATTTGAACGT   9300

CTTGATTATT CATAAATTGG ATACGTTCAT CATGATGTGA TAATTCGTCG GCATTTGTAA   9360

AACCTGTCTT TTTTTcAAGG CCTTCTAACA TTACTTTCAT CGGTACACCT TTAGGATCTG   9420

CTGATATCGC ATTCATCGTT TCTTTTTGAA TATCTTCAAT GACATAATGT TCTTCAAACG   9480

TAATACTTTT CATTTACTTC GCCTCCATAT TGTATTGCAT GTTTATTGCA TCTATTGCAG   9540

AAGCATTTTT TATATACCTC TAATTTCAAT GTTTGTAACA TAAAATTGAT CTACCAAGGC   9600

ATCTCTCCAT CGCCATTAAT AAATGTACCT GTTGGGCCAT CTGCACCAAT CGTTGCTAAT   9660

TGAATGATTG GCTTGATTCC TTCAGAAACG TGTTTGGAAT TATTACTAAA ATCACCAACT   9720

AAATCAGTAT TTGTAGCGCC TGGATCAGCA GCATTGATTT GCATGTTAGG TAATCCTTTA   9780

GCGTATTGTA GCGTTAGCAT TGTTACTGCC GATTAGACG AACAATAAGC TAATGAATTC   9840

ACTTTAGATT CAGCTGTTTC GGGGTTTGTA ACCATTCCAA ATGAACCTAA ACCACTTGAT   9900

ACGTTGACGA CAACAGGTTG TTCAGATTTT TCTAAGAGAG GGACGAATGT ATTCATCATT   9960

CGTACGATAC CGAATACATT CGTTTGATAT ACTTCTTCAA CGTCACGAGG TGTCAATTTG   10020

GAAGGTGCTG AAAATTGACC AGATATACCT GCATTGTTAA TGAGGATATC AAGACGGCCT   10080

TCTTTTTCAG CAATCATGTT ATAAGCATTT TTGACTGAGT AGTCACTTGT AACATCTAAT   10140

TGTACATAAT GAACACCTAA TTTTTGTGAT GCTTGTTGTC CTCTTACATC ATTCCGAGAA   10200

CCTATATAAA CTTTGTAACC CAATGCTTTA AGTGCCTCTG CACTTGCATA GCCTAACCCT   10260

TTATTGCCTC CTGTGATTAA CACAATTTTA GTCATTACGT CCCACCTCAT CTAAATAAAT   10320

GTTTAATAAA TAATTTCTGT ACGCTTCAAT TGAAATATGG CGATGCTCTA TTTGGAAGGC   10380

AAATACACTA GTTGATAATG ATTGCAACAG CATATCTGTT TTGAAtTCGT GTAAGTGTCG   10440

TCATCGCTTT TAAATAAGTC ATAATAAAAA TCAAATAATT CTTGATAAAA TGCGCTTTGG   10500

TAAAAACGTA ATTTATTGTT GCCTGCTTCA ATACATTGCA GTAGTGCCTT ATTATCGATT   10560

TTAAATTGTA AAAGATAATC TAACGACACT TGCATAACCT CATAATTAGA ATGATAGTCA   10620

TCTTTAATTT GCTTAAAATG AGTGATAAAA ATATCAAGGT CTCTTTGTAT GACGTAGTAG   10680

CATAAATCGC TTTTATCTTT GAAATGTCGA TACAATGTCC CCATACCGAT ACCTAGTTCT   10740

TTAGCAATAC GATTCATACT AATGTTTTCA ACGCCTTCTT CATCAAAAAG TTTGTGCGCT   10800

ATTTCTTCAA TTCGTTGCCT ATTCTCTTTT GCATCTTTTC GCATGATTAC ACCTACTTAA   10860

AATTCTCTAA AATTGACAAA CGGATAACTC TCCGTTTATT ATAAAACGTG TTAAGAAAGT   10920

TAGCAATGAA TTTGCAATAA CTATTAAATA TCATAAAAGA AAAGAGTGTT GATAATGTCT   10980
```

```
ACCTTATCGG TTCAAATGAT TGCTGAAAAA CTGAATGTCA CTACAGAAGA TGTGGAAAAA     11100

GTATTAGCTA TGACAGCGCC ACTAGGCATT TTTAGTCATC AATTACAACG ATTTATTCAT     11160

TTAGTATGGG ATGTCAGAGA TGTAATAAAC GACAATATTA AAGGAAATGG ACAAACACCA     11220

GAACCATATA CGTATTTAAA AGGTGAAAAA GAGGACTATT GGTTTTTAAG A             11271
```

(2) INFORMATION FOR SEQ ID NO: 12:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6261 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
CAACCCGTTC AGAACAAAAT AAAAACCGTA CAATTTTATC ATCTTAATGA TTATTGTACG       60

GAAAAACTTT TTTACATCAT ATCTGCATGT GCATAATCGA TATCGGTAAA TTTATTATAT      120

TGTTTCATAA AATGTAACTT AACTGTGCCT GTTGGACCGT TACGTTGCTT AGCAATGATA      180

ATTTCAATTT CACCGTTTTC ATCATTCGTT TGTGGCTCGA AACCACCATC ATCGTCATCA      240

TCTTCATCGC CGCCACGGTT ATAGTAATCA TCACGGTATA AGAATGCAAC GATATCGGCA      300

TCTTGCTCAA TCGAACCAGA TTCACGAATA TCACTCATCA TTGGACGTTT ATCTTGTCGT      360

TGTTCAACAC CACGAGATAA CTGACTTAAT GCGATAACTG GACATTTTAA TTCACGGGCT      420

AATGCTTTTA ATGTACGAGA GATTTCAGAA ACTTCCTGTT GTCTGTTATC GGACGCACGT      480

GAACCACTAC CTTGAATCAA CTGTAAGTAG TCAATCACAA TCATGTCTAA GCCATGTTCT      540

TGCTTTAATC GACGACATTT AGAACGTAAA TCATTAATTC GAATACCCGG TGTATCATCA      600

ATAAAAATCT TCGTACGTGA TAATTTACCT ACCGCTATAG TAAAACGACT CCAATCTTCC      660

TCAGTCATAG TACCCGTTCT TAAGCGGTTT GAGTCAACAT TTCCAGAACT ACAAATCATA      720

CGTGTGGCTA ACTGATCAGC ACCCATCTCT AGCGAGAAAA TACCAACTGT ATACATATCT      780

TCATGCGTTG CAACTTTTTG TGCAATATTA AGTGCGAACG CAGTCTTACC TACAGATGGA      840

CGCGCTGCAA GGATAATTAA ATCATTTCGG TTGAACCCTG CTGTCATTTG GTCTAAATCT      900

CGATATCCTG TAGGTATACC TGGTGTTTGA CCACTATTTT GATCAAGCTC TTCAGCTGTT      960

TCATACACTT GTCCTAAGAC GTCTCGAATG TCTTTAAAGC CATCGCTTTC ACGAGAAGAT     1020

GATAGCTCTA AAATTCGACG TTCTGCATCA CTTAAAATCG CATCTAGTTC AAGTTCATCA     1080

TTATATCCAT CATTGGCAAT ACTATCTGCA GTTTGAATCA ATCTACGTTT TAATGCATGC     1140
```

```
TCTGCAAGAT ATTGCGGGCC ACCCGCTTcA TTCAACGTAC CTTCCGTCGA TAATTGATCC    1260

ATCAATGTTA CAACATCAAT TTCTTTATTA TCTTCATTTA AGTGCATCAT TGCACGGAAA    1320

ATATGTTGAT GGGCACCCCT ATAAAACGAC TCAGGAAGCA AAACTTCCTG AGTAGTATTA    1380

ATCAATTCTG GATCTATAAT AATTGAACCT AAGACAGACT GTTCAGCTTC ATTGTTATGC    1440

GGCATTTGAT TTTGCTCATA CATTCTATCC ATGAATGGTT ACACCTCTTA TTTCAATCCA    1500

ACTTTATTGT TCAACTGTGT GTACGCGAAT TGTACCTTCA ACTTCTTTAT CTAATTTAAC    1560

AGGTACATTC GTATATCCTA GGGAATGAAT TCCATTTGGT AAATCCATTT TACGTTTATC    1620

AATTTTAATA TCATGTTGTG CTTTTAGTGC TTCGGCAATT TGTTTTGTAC TTACTGACCC    1680

AAACAATTTA CCACCTTCAC CAGTTTTTGC TGaTACTTCA ACTTCAATGT TTGATAACGT    1740

TTCTTTTAAT GCTTTAgCAT CTTCAATTTC TTGTTGGCGT TCTTGTTTTG CACGTTTTTT    1800

CTGTAACTCT AATTGTTTAA GGTTACCTGG TGTTGCTTCT ACAGCATAAT TCTTTTTCAA    1860

TAAGAAGTTA TTTGCATAAC CTACTGGTAC TTCTTTAACT TCACCTTTTT TACCTTTACC    1920

TTTACCTTTA ACATCTTGTG TAAAAATTAC TTTCATGCAT CTTCACTCCT ACTTAATTGT    1980

TCTGTAATTG CTTGTTGTAA TTGTGCTATC GCCTCTTCGA CTGTCACACC TTTAAGTTGT    2040

GTTGCCGCAT TGGTTAAATG TCCACCGCCA CCAAGTGCTT CCATTGTTAA CTGGACATTT    2100

ACTGAACCGA GTGAACGCGC AGATATACCA ATCAGATTAT CTTCACGTCT CGCAACAACA    2160

TATGATGCTT CAATACCTTC TAAACTTAAC AGTTCATCTG CTGCTTGTGC AACTGTTACT    2220

GGATGATAAA TTTTATCGTC TGAACCATGC GcAATGGCTA TGCCATTATC TTCAACTTTT    2280

ACAGTTCGAA TTAATTCAGA TCGATTAATG TAAGTATCCA CATCATCTTT TAAGAAATGT    2340

TGCGTTAAAA TCGTATCTGC ACCATGTGCA CGTAAATAAC TCGCTGCATC GAATGTTCTT    2400

GATCCTGTTC GTAATGTAAA GTTTCTTGTA TCTACAATAA TACCTGCATA CATCACTGTT    2460

GATTCAAGAC GTGTTAAACG TTGTTCTGTT GGTTGATATT CCAGTAACTC TGTTACCAAT    2520

TCAGCTGTCG AACTTGCGTA TGGTTCCATA TATATCAACA ATGGATTAGA GATGAAGCTT    2580

TCACCACGTC TATGATGATC GATAACAACT TTACGGTTTG CTTTATTTAA GACATTTTCA    2640

TCTAAAACCA GTTCCGGTTT ATGCGTATCA ACAATCACTA CGGTTGTCTT AGATGTCATC    2700

ATATCCCAAG CATCATCTGA TGTAATAAAT CGCTCTCTTA ACTCTGGCTT TTTATCTATT    2760

TCGTTCATCA CGCGTCGTAA TGTTGGATCA ATGTCAGTCT CATTTAATAC GATGTATGCT    2820

TCTAAATTAT TCATCATTGC AAATCTAGAC ACACCGATTG CTGCACCAAT TGCATCTAAG    2880

TCAGGACGTT TATGTCCCAT GATAATGACT TTGTCACCCT CTGCAAGGAT ATCTTTTAAC    2940
```

```
CCATAGAAAC GCACATTACC ATTAATACTT TTAATTGCAA CTTGGTCGCC ACCGCGTCCT   3060

AATGCTAAGT CTAGGCCTGA TTGTGATAAT TCACCTAAGT CGATTAAATT TTCAGTACCT   3120

TCACCAACAC CGATACTTAA TGTTAATTGG GCACGATAAC CAACACTTTT TTCACGTAAT   3180

TGACTCAAGA TATCAAATTT AGATTCTTCT AAGTCAGCTA ATATTTTTG ATTTAAATAG   3240

GCTACGAATT GATCGGAACT GTATCTTTTG AAAAATATAT TATACTCAGT TGCCCATCGA   3300

CTAATGACAC GCGTTACCAT TGAGTTGATT TCCGAACGCT GCGTATCATT CATATTTTGC   3360

GTAATCTCAT CGTAGTTATC TAAAAATAAT GTCGCAATGA TTGGTTTAGA ATTTTCATAT   3420

AGTTCATTTG TTTGTACTTG TTCAGTTATA TCAAAGAAAT AGAGGCAGTG ATCATTCTCA   3480

GAATAACGTA CTTGGAAATG ATACTGATTA TATTCTATTT CAACGGATTT CACTCTATCT   3540

AATTGCTTTA AAATGTTTGG AAATACTTCA TTTACAGATT CAGAAATGAC ATTCGCTTCC   3600

ATATGATCTG TCATAAATTG GTTAACCCAT TCGATGTGAT CATTTTCATC TAAAACAATG   3660

ATACCAATTG GTAAATGTTT GATTGCTTTA TTATTTGTTG TTGAAATTTG AGCACTCAAA   3720

CCATCTACAT AACTATCCAT TTTCATTAAA GCTTGTCTGA ATAAAATGAT GCTAACAATA   3780

ATCATCACGA CAAGAACGAT AGATGCAATT AGTGCTATAA GACTATTAAA GATAAACCAT   3840

ACACCCATTA AAACAATTGC TGTGATGATC ATGATGACAA ATGGTATTAG TAAAGCTTTC   3900

TTAGTGGACT GCCGATTCAT TATTCCACCT CTATTCACTT TTTAGAATTA TTTTTCATGA   3960

TTCGCTTCAA ATTCAAACTT AAATCGATAA CACCAAGTAG TCCTACAATA TGTGTCGTAG   4020

GTGTCAGTAT TGTACCGATA ACCAATAGTA AAATCGTTAC TGCATTCGGC AAACCTTTCG   4080

CTTTACCAAA GAAATGAATA ACACTTAAAC CTTGAATATA CATTACTAAT GATAACACAA   4140

GTTGGAAGTT TAAAAGAATG CTCTGGAACA CACTCGGTTG ACCTGTAAAT AATAAACATA   4200

TGATAACAAT AATGTATATC CATAATAAAA TACCGCTCAT TTGCCACGCG AAAAGTGGCT   4260

TAAATACAGG TGTAGCGATT TTAAATTTTC GTAAAATCGG AAATGTAACG ATTAAGTTAA   4320

TTAAGACGAT TAAAAATGTA ATGATAATGA TGAAACCTGG TAATTGAACG GTCGCTTGTC   4380

TAAACCCTTC TTCTAATATT TGGGTCATAT TCGCATCGGC ACCGCTCATC GTAATCGCTT   4440

CATGTAATGT TTGCTTGAAA GGTTTTACTA TGCTCGCTGA TGGTGGAATC CTTCCGAATG   4500

TTTGTAGTAA CATAAAAGCG ATTAATGAAA TTnArCTCAT CGCTACTGTT GTTACGTATA   4560

ATATTCTTTC TTTAGACGTT CTTTCTTTGA GCAATTGACC AATAATTAAA CTTGCAATTA   4620

AGACTAATAT GATGGCACTT AAAACGAAAG TATTACCTAA AACAGTTGTT ATAATTACTG   4680

TAATAAGTGC ACTAATCCCG AAAGATTGTA TTGATTTATT CCATAAAACG ATACCTGGTA   4740
```

```
CAAATACCAA CGCAATCGTT GCAATTATTG TTGCTTTAGG TTGTATTTTT GAAAACACAT    4860

AAGCCACTCC CATATTTTTA ACTATAGCTA TTATTTTAAC CTCTTTAATG AAAATTAACA    4920

ATTTATAGAT TGTATGCTTC TATTTCATTT AATTGAATAA TAACTTTCAT GTTTTATAAG    4980

TAATTAACAT ACTCATTTGA ATCGCTTTTG TGTGCTTTCA TTTTCAACAT GATTATTTAA    5040

TCCCACTACA TAGCAATCAA GCTTGATTTA GATTTACAAT ACATTTCCAC TCTCATGTAC    5100

TCTAGATGTT TTTGAATATG ATAACTGTGA TTTAGTGGCT TCATTCTTTG AAAATATATA    5160

TTATTACTTA CGCTTAAAAT GCTTTAAATT TAAGAAATGA TATAAGTTAG GTGCCCAGGT    5220

ACTAAAGTTT AGTAGGaATC CATCATGCCC AACATTATCA GGCACGAAGA AATGACGATG    5280

ATATTTAAAA CGTTCACCTA ATGCACGAAC TTGATCATCC GGATATAGCA AATCATCTAT    5340

GAACCCCATC GTTAACACTT TTGTTTCTAA ATTTTTAAAA ACATGCGTTA CGTCTGTGCG    5400

ACCTCGGTCA ATGTTGTGAC TATCCAATAC ATCTAGCAGT GTCAGATAAC AATTCAAATC    5460

AAAATGTTCT TTAAATTTAT TACCTTGATG TTGTTGGTAT GCGACTACTT CATCCGGCGT    5520

AAAACGTTCA TCATAACTTT TTGATGATCG ATATGTCAAA AAACCTAATT GGCGTGCAAT    5580

ACTTAGACCT TCCTTACCAC CAAGATGAAT GGCTTGCCTT GCAATTTCAT TGAAAGCTCT    5640

ACTATAAGAT GATGTTCGAC TTGTTGCAGC AAGGATAATG GCTTTATCTA CTTCAAACTG    5700

TTGATTGTAG AGTAGTTCCA TTGCTTGCAT ACCTCCAAGA CTTCCCCCTA TTAAAATATT    5760

AATCTTATCA TAACCAAGGG CTTGTATACC TCGTTCATTC GCTCTGACTA TATCTCTTAA    5820

TGTTAATTTT TTAGGAAAAT GAGGGTCGTT TAAAGGTGAA CTTGAACCGA AAGGACTACC    5880

AATAACATCA AATGTTAAAA ATTGATAATC GTGAATGGGT ATATATCCCC CATCAATAAT    5940

TTCTCGCCAC CAACCCGGAT AATCATCTGT TCCATATGTT AAATGATTGC CAGTTAATGC    6000

ATGACAAACT ACAACTAATG GTTGTCCATG ATAACCGACA TGCTCATATC TCAAACGCAA    6060

GTnATCTATG ACTTCCCCAG ATTCTGTAAT AAATTCCCCT AAATTTAAAG TATCTACTGT    6120

GTAATTTGTC ATTGTTCTTT CCTCCTTAAA CAAAAAAACT TCTCACCCTA TTGAAAAGTA    6180

AGAAGTCTTT ATACTTATCA TTCGAGTAAC TCGTTGGTTT TAGCACCGTG CTATAAAGTC    6240

GGTTGCTGAA GTATCACAGG G    6261
```

(2) INFORMATION FOR SEQ ID NO: 13:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1222 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
ATGCGATTAA CTCTGGAAAT ATCTTTTCCA TATTTACGTn TTAAATTATT CAGCAAATTC        60

ATACGAGaTT CATACTCGTT yAACACTTGT TCGTCGAATT CTGTATTAGC CATTTCATCA       120

TATAACTCAT GTTTTGCATC TTCTAAAATG TAGTAAAATT GATCAATATC TTCTTTTAAT       180

TTGTCATATT TGTTTGGAAC TATATCGTTT ATTGTTAACA AATGGTTGCT TAGTTCATAT       240

AAACGATCAG TGATAGCATT TTCATCCGTT AATGTCATAT ATGCGTTATT AAGCGCTAAG       300

CTTAATTTTT CAGAGTTTTG AATGCGTTTA ATATCTATTT CAAGTTGCTC TATTTCGCCT       360

TCTTTTAGAT GTGCTTCAGA CAATTCTTCT AATTGGAATT TCATTAAATC TAAACGCTGT       420

AGCAATGCTT GGTCTGCTGA TTCTAAATCT TCTAACTCTT GCTTTTTGGC TTTATAATTT       480

TGAAAAGTTT GGTGATATTT ATCCAACAAA TCTTGATAAC GTGATTCTGC GTAATTATCC       540

AATAATGTTA AATGGTATTT TTGTTTCAAC AAAGACTGCG TTTCATGTTG GCCATGAATA       600

TCTAATAATT CTTGCATAAC TTTTCGTAAA TCTTGTAAAG TAACTGTTTG ATTATTAATT       660

TTACAAAGAC TTTTACCAGA GCTGAAAATT TCCCGTTTAA CTAATAAAAA ATCTTCATCT       720

ACATCAATAT CCATATTTTT CAATATATGT ATAGCATCTT TACTCTCGTC AATATCAAAT       780

ATACCTTCGA TGACAGCCTT TTTTTCACCA TGTCTTACAA AATCAGATGA AGCTCTCATT       840

CCAATTAATT GTCCAATTGC ATCTATAATA ATTGACTTAC CTGAACCCGT TTCACCACTT       900

AAAACAGTTA AACCATCAGA AAATTGAATT TCTAATTCTT CAATAATAGC AAATTGCTTG       960

ATTGATAAGG TTTGTAACAT AAACTCATCG CATCCTTATA ACAAATTGAA AATTCTTGAC      1020

TTGATTTCAT CACTTGCCTC TTTGCTTCGA CAAATAATTA AACAAGTATC ATCACCACAA      1080

ATTGTGCCTA GTACTTCTTC CCAATTGATT TGGTCTAATA TAGCTCCAAT AGATTGTGCA      1140

TTACeAGGTA TGTTTTTAGA ACAAGTAAAT TATCAGTACC ATCTATATTA ACAAAGGAAT      1200

CCATTAAATA ACGTCCCAAT TT                                             1222
```

(2) INFORMATION FOR SEQ ID NO: 14:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1021 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
TTTGTTATTA TTACnTnAAA TAATTGCATT ACTTTTTACT GATGGTACAA CTTTCCATCC        60
```

```
TTCTTTTGGC ACGACATAAT TGTCTTTATC TTGAACTAAA TATCCGCCAG ATACTGAAAC      180

AAACTCTTCT TCGTTACTGT CTATAGTCAT ATCAATTTCT AATAATCTTA CATTCTTCTT      240

TTGTTTTAAA ATATCTAATG CTTCATCTGT AAATTTTGGT GCAATAATGA CTTCCAAAAA      300

GATACTATGC AATTGCTCTG CTAACTCAGG TGTTACAGCT CGGTTTAATG CAACAATTCC      360

ACCAAATATT GATTGACTAT CCGCTTCATA CGCATGTTGA AATGCTTGTT CTATCGTGTC      420

ACCGATACCA ACACCACATG GATTCATGTG TTTAACCGCA ACTGTAGCAG GTGTATCAAA      480

CTTTTTAACT AAAGCTAGTG TAGCATCTGC ATCTTTAATA TTGTTATAGC TTAATTGTTT      540

CCCATGTAAT TGTTTAGCGC CTGCAATCGT GTGCTTAGCA TTCGAAGTTC TCACAAAATA      600

CGCTGATTGT TGTGGATTTT CTCCATATCT TAAAGTTTCT TTATCCCCTT TAAAGAAACG      660

TACAATCGCT TCATCATATT CTGCAGTATG CTCAAAAACT TTAATCATTA ATGATTGTCT      720

ATATGACTCA TCTAACGAAT CGTTTCTTAA TCGCGTCAAT ACTTCTTGAT AATCTGCCGG      780

ATGTACAATT GTTGTTACAT GTTTATAGTT TTTAGCTGCA GCACGTAACA TTGTTGGACC      840

ACCAATATCA ATATTTTCAA TTGCTTCGTC CATCGTCACA TCAGGGTTTG CAACAGTTTG      900

TTGGAATGGA TATAAATTAA CTACTACCAT ATCAATTAAA TCTATATGTT GTTCTGATAA      960

TTCATTTAAA TGCTGCGGTT TATTTCGATC AGCTAAAATG CCACCATGAA CAGCCGGATG     1020

T                                                                    1021
```

(2) INFORMATION FOR SEQ ID NO: 15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3759 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
TCATTCACTC CTAAATTGTT ATTACACTAT TACACaTAGC TAATCATCAA TGTGAAATCA       60

CCTTCAAAGA CACTATCCAA ATCTTCAGAA GTCAAAATAA AGTTTGTACC AGTAGTCAGT      120

TTGAAAATTT CACCATCGAC AATCATTTGC CCTTCGCCTT CCAACACTGT AACTAAACAG      180

AACTCTCTAG GCTTCATATA ATTTAACGTG CCAGAAATTT CCCATTTAAC CAATGTAAAG      240

AAATCATTCG ATACAATGTG TGTACACTTA TGGTTTTCAA TAATTTCGCT TTCAGGCAAA      300

ATATTAGGTA ATGGTGCATT GTACTGAATA ACGTCTAAAG CTTTTTCAAT ATTTAACGGT      360

CTATCATTAT ATTGATTATC TTGACGATTG AAATCATAAA GTCTATATGT AATGTCTGAC      420
```

```
ATAAAAtAGa ATTCyCCAGG kTTTACtTTA AtatATCyAA gTAtCGaCtC tATCGTTCCG    540

TGTTGAACAT GATTCGCAAC TTCTTCTCTA GACTCTGCTA ATGTCCCtAT AACTATTTCT    600

GCATCTTCTT CTGCATCTAT AATATACCAA CATTCAGATT TGCCATATTG CCCgTTTTCA    660

TGCTCATAAG CATAAGAATT ATCAGGGTGC ACATGAATAG AAAGTGATTC TCTTGCATCC    720

ACTATTTTAG TTAGAAGCGG AAAATCTTTG CTTGGGAAAT CACCAAACAA TTCACGATGT    780

TCTGACCAAA TACGGTCTAA TGTTTGACCT TGATATGGTC CATTAATAAT CTCGCTCGTA    840

CCATTTGGAT GTGCTGACAC ACACCAACAT TCCCCCAGTT GTATCATTGT CTAATTGATA    900

TCCAAACTCA CTTAGACGTT GACCGCCCCA TAATTTTGTT TTTAAAATTG GTTGTAAAAA    960

TAATGGCATT GTTGCACCTC CATTGTGATT AAGTAAGCAA TAGAACTCTG ATGTTGTTGT   1020

TCCATTATAT TTTGATTTTG TTCTCATTTA CATCGTATTA TTAACTTCCA CATTTCAAAT   1080

TAACTATTAG TGATTGTACC ATATTTACTA ACATTGCAGT ACTGCCAATT AAAAGnGCTT   1140

CACTTAAATT TACAGTACTT TAACATTTTC AAAAATTTAT AGCATAGAGA TTATATCTCT   1200

CTTACATTTG TACATATTTC CCTTTAAATT TACTCGCCCA TTATACCAAT TAATAaACAA   1260

CTTTAATAGT TGTGCCATAC ATTGTTCAAA TTCTTTGTAA AACGCATAGA CAATACGTAC   1320

TTATTCATAC TTATAATTCA TCATTTTCAA AAAATAACGA GTTACGAAAA AGTAACCCGC   1380

TTCAAATCAT ATTTACTATC CTTATTAATC CGTTTCATTT TCAAATTGAG TTAAAGCATC   1440

TTTAATGTCC TGATCACCAC TAATAATTTG AAACTCTTGG TGATTAAAAT GATTGGATGT   1500

GACAATTTCT TTTAATACTG TCGCAACATC TTCTCTAGGA ATTTCACCTT TACCATCAAA   1560

ATATTGTGCA GCTTCTATCT TTCCAGATCC TGCTGCATTT GTAAGTGCCC CTGGATGTAA   1620

AATTGTATAA TTCAAACCTG nAACGTCTTA AATAGTCATC AGCGTAATGT TTAGCTATTG   1680

TATATGGCTT TAAATCACCG CTATCATCAA AAGCCTGACG TCTCGAATCA TATGTTGAAA   1740

CCATGACATA GTGTTTAATA TTGGCCTCTT TACTCGCAAT CATTGATTTA ACAGCACCAT   1800

CTAAATCGAC AATAATTGTT TTATCTGCAC CCGTGTTCCC TCCAGAACCT ACTGAAAAGA   1860

TAACTTTATC GAATGGTTTA AACGTCTCAG TTAAAGTCTC TATTGAATCA TTTTCAACAT   1920

CAACAAGAAT TGCTTTCATA CCTTGTGATT TTAACGCATT AAGTTGATCT GATTGCCTAA   1980

CACCAGCAGT AAATGGTACA TTTTCTTTTG CTAATTGTTG CACTAGTAAC GAACCTACAC   2040

CGCCATTAGC ACCTATAACC AAAATATTCA TTTACAACAC TCTCCTATkT ATTATTCTCT   2100

ATGCCATACC ACTTTATGAG ATATGTAAAA CTTGTTACAA CTATAAAAAT CAATTGACAT   2160

ACTACTGGGA ACGTATTAAA TTAATATATG AACAAATATT CATATGAAAG GATTGTCATA   2220
```

```
tCaAGGCATT AGcGATTACA ATCGAATACG TATCaTGGAA TTGTTATCaG TCAGCGAAgC        2340

AAGTGTTGGT CACATTtCAC ATCAATTGAA TTTATCTCAA TCAAATGTCT CGCACCAATT        2400

AAAATTACTT AAAAGTGTGC ATCTTGTGAA AGCAAAACGA CAAGGCCAAT CAATGATTTA        2460

TTCATTAGAT GACATCCACG TAGCAACTAT GTTAAAGCAA GCCATACATC ACGCGAATCA        2520

TCCTAAAGAA AGTGGGTTAT AATATGTCTC ATTCACATCA TCATCATGAC CATATGCATA        2580

GTCATGTAAC TACAAATAAT AAGAAAGTAT TGTTTATATC GTTTTTAATA ATCGGTCTAT        2640

ATATGTTTAT CGAAATCATC GGCGGTCTCC TTGCTAACAG CTTGGCATTA CTATCTGACG        2700

GTATCCATAT GTTTAGCGAC ACATTCTCAT TAGGTGTTGC ACTTGTCGCA TTTATTTATG        2760

CTGAAAAGAA TGCCACAACT ACAAAAACAT TTGGTTATAA ACGTTTCGAA GTACTCGCAG        2820

CGTTATTTAA CGGTGTAACG CTTTTTGTAA TAAGTATTTT GATTGTTTTT GAAGCGATTA        2880

AACGTTTCTT TGTTCCTTCT GAAGTTCAAT CAAAAGAAAT GTTAATCATT AGTATTATCG        2940

GTTTAATTGT CAATATCGTT GTTGCATTCT TTATGTTTAA AGGCGGCGAC ACTTCACACA        3000

ATTTAAATAT GCGTGGTGCT TTTCTACATG TTATCGGAGA CTTATTAGGT TCAGTTGGCG        3060

CCATTACTGC AGCTAkTTTA ATTTGGGCAT TTGGATGGAC AATCGCCGAT CCTATCGCAA        3120

GTATTTTAGT TTCCGTTATT ATTTTAAAAA GTGCTTGGGG TATCACAAAA TCTTCAATTA        3180

ACATTTTAAT GGaAGGCACA CCAAGTGATG TTGATATAGA TGAAGTTATA ACTACTATTA        3240

AAAAGGATTC ACGAATACAA AGTGTGCATG ATTGCCATGT TTGGACAATT TCAAATGATA        3300

TGAATGCATT GAGTTGTCAT GTTGTTGTAG ACCATACATT GACAATGAAA GAATGTGAAT        3360

TATTATTAGA AAaCATTGAG CATGATTTAT TACATTTAAA TATTCACCAT ATGACTATTC        3420

AATTAGAAAC GCCTAATCAC AAACATGATG AATCGATTAT ATGTTCAGGA ACACATAGTC        3480

ATTCACATAA CCATCATGCT CATCATCACG CGCATGTACA TTAATAATTT TAACCTACTG        3540

CCATTGCATC GATTAAACTT TTCAATGGCA GTAGGTTTTT TATGTCTTTA TGGCGACTTG        3600

TTTGGTCTTT GATGATGCAA TGTTTATTAA CAAATTTTCA ACTATTATTT CTTACATTAG        3660

TCATATTTTT GACAATTTAC TATTATAATT CTCTAACTTT AGTCACTTTA ATTAATTTTT        3720

ATTAGATATT AATATGAAAA TAACGTGTTT TTTGTTATT                              3759
```

(2) INFORMATION FOR SEQ ID NO: 16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 13086 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
TAATTATCGC GCATAACAAA ACATTAGCAG GACAATTATA TAGTGAGTTT AAAGAATTTT        60
TTCCTGAAAA CAGGGTGGAA TACTTTGTAA GTtACTATGA TTATTATCAn CCAGAGGCAT       120
ACGTACCGTC TACTGACACT TTTATTGAAA nAGATGCCTC AATCAnTGAT GAAATTGATC       180
AACTACGACA TTCTGCTACA AGTGCATTAT TTGAACGCGA TGATGTAATT ATTATTGCTA       240
GTGTAAGTTG TATATATGGT TTAGGTAATC CTGAAGAATA TAAAGATTTA GTAGTAAGTG       300
TTCGAGTTGG TATGGAAATG GATAGAAGTG AATTACTTAG AAAACTTGTc AGATGTGCAA       360
TATACACGAA ATGACATCgA TTTcCAACGA GGAACGTTTC GAGTGCGTGG TGATGTAGTG       420
GAAATATTCC CAGCCTCTAA AGAAGAACTT TGTATAAGGG TTGAGTTTTT CGGCGATGAG       480
ATTGACCGTA TCCGAGAAGT TAACTACCTA ACAGGTGAAG TGTTGAAAGA AAGAGAACAT       540
TTTGCGATAT TCCCAGCTTC TCACTTCGTA ACACGTGAAG AAAAGTTGAA AGTTGCGATT       600
GAACGTATTG AAAAAGAATT GGAAGAACGA TTGAAAGAAT TACGAGATGA GAATAAATTA       660
CTAGAAGCGC AAAGGTTAGA ACAGCGTACC AACTATGATT TAGAAATGAT GCGAGAGATG       720
GGATTCTGTT CAGGAATTGA AAACTATTCC GTACATTTAA CTTTGCGACC ACTGGGTTCG       780
ACACCATATA CTTTATTGGA TTACTTTGGC GATGATTGGT TAGTAATGAT TGATGAATCA       840
CATGTGACAT TACCGCAAGT TCGAGGCATG TATAACGGAG ACAGAGCGCG TAAACAAGTT       900
TTGGTGGATC ATGGGTTTAG ATTACCGAGT GCATTAGATA ACCGTCCACT TAAATTTGAA       960
GAATTTGAAG mAAAGACAAA ACAACTTGTG TATGTATCTG CAACGCCTGG ACCATACGAA      1020
ATTGAACATA CGGATAAGAT GGTTGAACAA ATTATTCGTC CTACTGGTTT ACTGGATCCT      1080
AAGATTGAGG TTAGACCTAC TGAAAATCAA ATTGACGATT TATTAAGTGA AATTCAAACA      1140
AGAGTgAGCG TAATGAACGC GTACTTGTTA CAACGCTCAC TAAAAAGATG AGTGAAGATT      1200
aACCACATAC ATGAAAGAaG CGGGTATTAA aGTtAATTAT CTGCATTCAG AAATCAAGAC      1260
ATTAGAACGA ATTGAAATAA TTAGAGACTT ACGAATGGGT ACATATGATG TTATCGTAGG      1320
TATTAATTTA TTAAGAGAGG GTATTGATAT ACCAGAAGTT TCTCTAGTTG TCATATTAGA      1380
TGCAGATAAA GAAGGGTTTT TACGTTCTAA CCGCTCATTA ATTCAAaCAA TAGGTAGAgC      1440
TGCGCGTAAC GATAAaGGTG AAGTCATTAT GTATGCCGAT AAAATGACTG ATTCGATGAA      1500
GTATGCAATT GATGAGACAC AACGTCGTCG AGAAATACAG ATGAAACATA ATGAAAAACA      1560
TGGTATTACA CCTAAAACAA TTAATAAAAA AATACATGAT TTAATTAGTG CTACTGTTGA      1620
AAATGACGAA AATAATGACA AAGCACAAAC TGTGATACCT AAGAAGATGA CGAAAAAAGA      1680
```

```
TTTCGAGAAA GCTACAGAAT TAAGAGATAT GTTATTTGAA TTAAAAGCAG AAGGGTGACA    1800

AGTAAATGAA AGAACCATCC ATAGTAGTAA AAGGTGCTCG TGCGCATAAC TTGAAAGATA    1860

TTGATATCGA ACTACCTAAA AaTAAATTAA TTGTTATGAC AGGTTTATCT GGGTCAGGTA    1920

AATCGTCATT AGCATTCGAT ACTATATATG CTGAAGGACA ACGACGTTAT GTTGAATCAT    1980

TAAGTGCCTA TGCGCGTCAA TTTTTAGGCC AAATGGACAA ACCAGATGTT GATACAATTG    2040

AAGGATTATC GCCAGCAATT TCAATAGATC AAAAAACAAC AAGTAAAAAT CCAAGATCAA    2100

CTGTAGCAAC AGTAACAGAA ATATATGATT ATATACGTTT GTTATATGCA CGTGTTGGTA    2160

AACCTTACTG TCCAAATCAC AATATAGAAA TTGAATCGCA AACAGTACAA CAAATGGTTG    2220

ACCGCATTAT GGAATTAGAG GCACGTACAA AGATTCAATT ATTAGCACCT GTCATCGCTC    2280

ATCGTAAAGG TAGTCATGAA AAGCTAATCG AAGATATTGG TAAAAAAGGT TATGTACGTT    2340

TAAGAATCGA TGGCGAAATT GTTGATGTAA ATGATGTACC TACTTTAGAT AAGAACAAGA    2400

ATCATACAAT AGAAGTTGTT GTAGACCGAT TAGTTGTTAA AGATGGAATT GAAACACGAC    2460

TAGCTGACTC TATAGAAACT GCCTTAGAGC TTTCAGAAGG ACAATTAACA GTCGATGTCA    2520

TTGACGGGGA AGACCTTAAG TTTTCAGAAA GCCATGCTTG TCCTATATGT GGATTTTCAA    2580

TCGGAGAGTT AGAACCAAGA ATGTTTAGCT TTAACAGTCC TTTTGGTGCT TGTCCGACAT    2640

GTGATGGCTT AGGCCAAAAG TTAACAGTCG ATGTAGACTT GGTTGTTCCC GACAAAGATA    2700

AGACGCTAAA CGAAGGTGCA ATAGAACCTT GGATACCGAC GAGTTCTGAT TTTTATCCAA    2760

CATTGTTAAA ACGTGTTTGT GAAGTTTATA AAATCAATAT GGATAAACCT TTTAAAAAGT    2820

TAACAGAACG TCAACGTGAT ATTTTATTGT ATGGTTCTGG TGACAAAGAA ATTGAATTTA    2880

CATTTACACA ACGTCAAGGT GGTACTAGAA AACGAACAAT GGTTTTCGAG GGTGTAGTTC    2940

CTAATATAAG TAGACGATTC CATGAATCTC CTTCAGAATA TACACGTGAA ATGATGAGTA    3000

AATATATGAC TGAACTACCT TGCGAAACTT GTCATGGAAA GCGATTGAGT CGTGAAGCkT    3060

TATCTGTTTA TGTAGGTGGT TTAAATATTG GTGAAGTAGT CGAATATTCA ATCAGTCAAG    3120

CGCTGAACTA TTATAAAAAC ATTGATTTGT CAGAACAAGA TCAAGCGATT GCAAATCAAA    3180

TATTGAAAGA AATTATTTCC CGACTCACTT TTTTAAATAA TGTGGGACTT GAATATTTAA    3240

CGTTAAACAG AGCTTCAGGT ACACTTTCAG GTGGTGAAGC ACAACGTATT CGATTAGCAA    3300

CGCAAATTGG GTCGCGTTTG ACTGGTGTCT TATATGTATT AGATGAGCCA TCAATTGGAC    3360

TGCATCAAAG AGATAATGAT CGATTAATTA ATACACTTAA AGAAATGAGA GATTTAGGAA    3420

ATACTTTAAT TGTAGTTGAA CACGATGATG ATACAATGCG TGCGGCTGAT TACTTAGTGG    3480
```

259

```
AGGTAATGAA AGATAAAAAA TCATTAACAG GACAATACTT GAGTGGTAAG AAACGTATTG    3600

AAGTACCTGA ATATCGCAGA CCGGCTTCAG ATCGTAAAAT TTCTATACGT GGAGCTAGAA    3660

GCAACAATCT TAAAGGGGTT GATGTGGACA TACCACTATC AATCATGACG GTTGTTACAG    3720

GTGTATCAGG TTCTGGTAAA AGCTCATTAG TAAATGAAGT ATTATACAAA TCATTAGCTC    3780

AAAAAATTAA TAAATCTAAA GTAAAGCCAG GATTGTACGA TAAGATTGAA GGTATTGATC    3840

AACTTGATAA AATTATTGAT ATTGATCAAT CACCAATAGG TAGAACGCCA CGCTCTAATC    3900

CAGCAACATA TACTGGTGTG TTTGATGATA TACGTGATGT GTTTGCGCAA ACAAATGAAG    3960

CTAAAATTCG AGGATATCAA AAAGGGCGTT TTAGTTTTAA TGTAAAAGGT GGACGCTGTG    4020

AAgcTTGTAA AGGTGACGGT ATTATTAAAA TTGAAATGCA TTTTTTACCT GATGTTTATG    4080

TTCCTTGTGA AGTGTGTGAT GGTAAACGAT ATAATCGTGA GACACTAGAG GTTACTTACA    4140

AAGGTAAAAA TATTGCTGAC ATTTTAGAAA TGACTGTTGA AGAAGCAACA CAATTTTTTG    4200

AAAATATTCC TAAGATTAAG CGCAAGTTAC AAACACTAGT TGATGTTGGT CTTGGATACG    4260

TCACATTAGG TCAACAAGCT ACAACGTTAT CAGGTGGTGA GGCTCAACGT GTGAaACTTG    4320

CATCTGAACT TCATAAACGT TCAACTGGTA AATCTATTTA TATCCTAGAT GAACCGACAA    4380

CAGGGTTACA TGTTGACGAT ATTAGTAGAT TATTAAAAGT ATTAAACCGA TTAGTTGAAA    4440

ATGGTGATAC TGTTGTAATT ATTGAACATA ACCTAGATGT TATCAAAACA GCAGACTATA    4500

TTATAGACTT AGGTCCTGAA GGTGGTAGTG GCGGTGGTAC TATTGTTGCG ACTGGCACAC    4560

CCGAAGATAT TGCTCAGACA AAGTCATCAT ATACAGGAAA GTATTTAAAA GAAGTACTTG    4620

AACGAGATAA ACAAAATACT GAAGATAAAT AAGATTAAAA GAAGTGAAGG ATGTTATAAA    4680

TTTATCCTTC GCTTCTTTTT ATTAATTTAG TAATGAATAG TAGAAAGAAA AGATGCGTAA    4740

AAAGAATTAT GTTAAGATAG GGTCAATCTA GAGTAGTTAA ACATAAATCG AACTGGGAGT    4800

GGGACAGAAA TGATAAAGAA TCACTAATGA TTTATTATGT AGTGGTTCTT TGTCATTAGC    4860

CACAGCTATT GTGTACTTAA AAATAGGaat GCaTgAGTGC AACTCATGCA TAAGaAATAC    4920

TAATTTCTAA AGAAAAAGTA TTTCTTTATG TTGGGGCCCC GCCAACTTGC ATTGTTTGTA    4980

GAATTTCTTT TCGAAATTCT TTATGTTGGG GCCCCGCCAA CTTGCATTGT TTGTAGAATT    5040

TCTTTTCGAA ATTCTTTATG TTGGGGCCCC GCCAACTAAT TCCAATATAT CATTGTAGAG    5100

CTTAGGTCAT TGATTTTTGG CTCGGACTTT TATGGCGATA TGAACCATGT AAATTAAGCA    5160

AGCAATAAAT TAATGATTGA TATTGACTTG TAAAATAATA ACAATAATGA ACAATTAATA    5220

TTTATTTTAG CTTTTCAATG TAGATTGGTG TTATATTTTT GATATGATAA GAAGAGATGT    5280
```

260

```
ACATTAAAGT TAGATTTAAT CGCTGGTGAA GAAGGACTAT CGAAGCCAAT TAAAAATGCT    5400

GATATATCAA GACCGGGCTT AGAGATGGCA GGTTATTTTT CACATTATGC GTCAGATAGA    5460

ATACAACTAT TAGGAACAAC GGAACTATCG TTTTACAATT TATTACCAGA TAAGGATCGC    5520

GCAGGTCGTA TGCGTAAACT ATGCAGACCA GAAACGCCTG CAATTATTGT GACACGTGGA    5580

TTGCAGCCAC CAGAAGAATT AGTTGAAGCT GCAAAAGAAT TAAATACCCC ACTTATAGTT    5640

GCTAAAGATG CGACTACAAG TTTAATGAGT CGCTTAACAA CGTTTTTAGA GCATGCACTT    5700

GCAAAGACGA CATCTTTACA TGGTGTTTTA GTAGATGTTT ACGGTGTTGG TGTACTAATT    5760

ACCGGTGATT CAGGAATAGG TAAAAGTGAG ACTGCGTTGG AATTAGTTAA ACGTGGGCAT    5820

AGATTAGTAG CAGATGATAA TGTAGAAATA CGTCAAATTA ATAAAGATGA ACTAATAGGG    5880

AAACCACCAA AGTTAATAGA ACATCTATTA GAAATACGTG GACTAGGTAT TATCAATGTT    5940

ATGACTTTAT TTGGCGCGGG TTCAATATTA ACTGAAAAAC GAATTAGATT AAATATTAAT    6000

TTGGAAAACT GGAACAAGCA AAAGTTATAT GACCGCGTAG GTCTTAATGA AGAGACGCTA    6060

AGTATTTTAG ATACTGAAAT CACTAAAAAA ACAATACCTG TAAGACCTGG TAGAAATGTT    6120

GCGGTAATTA TTGAGGTCGC TGCAATGAAC TATCGATTAA ATATCATGGG CATTAACACG    6180

GCCGAAGAAT TTAGTGAAAG ATTAAATGAA GAAATTATCA AGAACAGTCA TAAGAGTGAG    6240

GAGTAGGTTG AATGGGTATT GTATTTAACT ATATAGATCC TGTGGCATTT AACTTAGGAC    6300

CACTGAGTGT ACGATGGTAT GGAATTATCA TTGCTGTCGG AATATTACTT GGTTACTTTG    6360

TTgCACAACG TGCACTAGTT AAAGCAGGAT TACATAAAGA TACTTTAGTA GATATTATTT    6420

TTTATAGTGC ACTATTTGGA TTTATCGCGG CACGAATCTA TTTTGTGATT TTCCAATGGC    6480

CATATTACGC GGAAAATCCA AGTGAAATTA TTAAAATATG GCATGGTGGA ATAGCAATAC    6540

ATGGTGGTTT AATAGGTGGC TTTATTGCTG GTGTTATTGT ATGTAAAGTG AAAAATTTAA    6600

ACCCATTTCA AATTGGTGAT ATCGTTGCGC CAAGTATAAT TTTAGCGCAA GGAATTGGAC    6660

GCTGGGGTAA CTTTATGAAT CACGAGGCAC ATGGTGGATC GGTGTCACGC GCTTTTTTAG    6720

AACAATTACA TTTGCCTAAT TTTATAATAG AAAATATGTA TATTAACGGC CAATATTATC    6780

ATCCAACATT CTTATATGAA TCCATTTGGG ATGTCGCTGG ATTTATTATC TTAGTTAATA    6840

TTCGTAAACA TTTAAAATTA GGAGAAACAT TCTTTTTATA TTTAACTTGG TATTCAATTG    6900

GTCGATTCTT TATAGAAGGA TTACGTACAG ATAGCTTAAT GCTCACAAGT AATATTAGAG    6960

TTGCACAATT AGTATCAATT CTTTTAATTT TAATAAGTAT AAGTTTAATT GTATATAGAA    7020

GGATTAAGTA TAATCCACCG TTGTATAGCA AAGTTGGGGC GCTTCCATGG CCAACAAAAA    7080
```

```
TTATGGCGTG TATACCGTCT TGTTAAATTT TCGAAAGTTT TTAAGAATGT AATTATCATT    7200

GAATTTTCGA AATTTATTCC AAGTATGGTA CTGAAAAGAC ATATATATAA ACAACTTTTA    7260

AATATTAATA TCGGTAATCA ATCGTCGATA GCTTATAAAG TAATGTTAGA TATTTTTTAC    7320

CCAGAACTGA TTACGATTGG TAGTAACAGT GTTATTGGTT ACAATGTAAC AATTTTGACG    7380

CATGAAGCAT TAGTTGATGA ATTTCGTTAT GGACCAGTGA CGATAGGATC TAACACTTTG    7440

ATTGGTGCAA ATGCTACCAT TTTACCCGGT ATAACGATTG GTGACAATGT AAAAGTTGCA    7500

GCTGGTACGG TTGTTTCAAA AGATATACCG GATAATGGAT TTGCATATGG CAACCCTATG    7560

TATATAAAAA TGATTAGGAG GTGACAATTT TATGGCGCAA AAGAATAATA ATGTAATTCC    7620

AATGACTTTT GATGATGCAT TTTATCGTAA AATGGCTAAA CAGAAGTTTA AACAAAGAGA    7680

ATATAAACGA GCTGCTGAAT ACTTTGAAAA AGTGTTAGAA TTGTCACCTG ATGATCTGGA    7740

AATTCAAATT GATTATGCAC AATGTCTAGT GCAACTTGGT ATTGCTAAAA AAGCAGAACA    7800

TTTATTTTAT GACAATATTA TTTATAATAG GCATCTAGAA GATAGCTTTT ATGAATTGAG    7860

TCAGCTCAAC ATTGAAGTTA ACGAACCAAA CAAGGCATTC TTGTTTGGTA TTAATTATGT    7920

TATTGTTAGC GACGACCAAG ATTATAGAGA TGAATTAGAT CAAATGTTTG ATGTGAAATA    7980

TCAAAGTGAA GAACAAATTG AACTTGAAGC TCAATTGTTT GTAGTTCAAA TACTATTCCA    8040

ATATCTTTTT TCTCAAGGTC GATTAAAAGA TGCAAAGAAT TATGTCTTAC ATCAACCACA    8100

AGAAGTTCAA GATCATCGTG TAGTACGTAA TTTATTGGCA ATGTGTTATT TATATCTCGG    8160

TGAATATGAT ACgGCTAAAG CATTGTACGA aGCACtATTA CAAGAGGATA GTACaGATAT    8220

ATATGCATTA TGCCATTATA CTTTGCTACT TTATAACACT AAGGAAAATG AACAATATCA    8280

AAAATATTTA AAAATATTAA ACAAAGTTGT ACCTATGAAT GACGATGAAA GTTTTAAATT    8340

AGGTATTGTA TTAAGTTATT TAAAGCAGTA TCGTGCATCA CAACAATTGT TGTACCCTTT    8400

ATATAAAAAA GGGAAATTTT TATCAATTCA AATGTACAAT GCTTTAGCAT ATAATTATTA    8460

TTATTTAGGT GAAGAAGACG AAAGTCATTA CTACTGGGAT AAATTGAAGC AAATTTCTAA    8520

AGTGGAAATT GGACATGCGC CTTGGGTAAT TGAAAATAGC AAAGAAGTTT TTGACCAACA    8580

TATTTTGCCA TTACTTCAAA GTGATGACAG TCATTATCGT TTATATGGTA TTTTTTTATT    8640

GGATCAATTA AATGGTAAAG AAATTGTGAT GACGGAAAGT ATTTGGCAGG TTTTGGAAAA    8700

TCTAAATAAT TATGAGAAAT TGTATTTAAC GTATTTAGTT CAAGGTTTAA CGCTCAATAA    8760

ATTAGACTTC ATTCATCGCG GCTTATTAAC GCTTTACCAT AATGAATTAT TTGTAAGTGA    8820

AAATGATGTA ATGGTTGCAT GGATTAATCA AGGTGAACTC ATAATTGCTG AAAAAGTAGA    8880
```

```
TCGAAACGTT ACAAAGAAGC AAATTACAAC ATGGTTAGGC ATAACACAAT ATAAACTGAA      9000

CAAAATGATT GAATTTCTCT TGAGCATATA GATTTATGAA AAGTTAGATT TATTATATAA      9060

TGCGCATAAT GATTAATAAT GAGGAGGCGT TAATAAAATG ACTGAAATAG ATTTTGATAT      9120

AGCAATTATC GGTGCAGGTC CAGCTGGTAT GACTGCTGCA GTATACGCAT CACGTGCTAA      9180

TTTAAAAACA GTTATGATTG AAAGAGGTAT TCCAGGCGGT CAAATGGCTA ATACAGAAGA      9240

AGTAGAGAAC TTCCCTGGTT TCGAAATGAT TACAGGTCCA GATTTATCTA CAAAAATGTT      9300

TGAACACGCT AAAAAGTTTG GTGCAGTTTA TCAATATGGA GATATTAAAT CTGTAGAAGA      9360

TAAAGGCGAA TATAAAGTGA TTAACTTTGG TAATAAAGAA TTAACAGCGA AAGCGGTTAT      9420

TATTGCTACA GGTGCAGAAT ACAAGAAAAT TGGTGTTCCG GGTGAACAAG AACTTGGTGG      9480

ACGCGGTGTA AGTTATTGTG CAGTATGTGA TGGTGCATTC TTTAAAAATA AACGCCTATT      9540

CGTTATCGGT GGTGGTGATT CAGCAGTAGA AGAGGGAACA TTCTTAACTA AATTTGCTGA      9600

CAAAGTAACA ATCGTTCACC GTCGTGATGA GTTACGTGCA CAGCGTATTT TACAAGATAG      9660

AGCATTCAAA AATGATAAAA TCGACTTTAT TTGGAGTCAT ACTTTGAAAT CAATTAATGA      9720

AAAAGACGGC AAAGTGGGTT CTGTGACATT AACGTCTACA AAAGATGGTT CAGAAGAAAC      9780

ACACGAGGCT GATGGTGTAT TCATCTATAT TGGTATGAAA CCATTAACAG CGCCATTTAA      9840

AGACTTAGGT ATTACAAATG ATGTTGGTTA TATTGTAACA AAAGATGATA TGACAACATC      9900

AGTACCAGGT ATTTTTGCAG CAGGAGATGT TCGCGACAAA GGTTACGCC AAATTGTCAC      9960

TGCTACTGGC GATGGTAGTA TTGCAGCGCA AAGTGCAGCG GAATATATTG AACATTTAAA     10020

CGATCAAGCT TAATTCGAAG TCGAATTAAG ATGTTGAGCT GTAAATTATT TGGATATTTA     10080

TTTTAATAGT GTCATCACAG CGTTAAAATA ATGTCTTACT TTTAAATTAA AGCAAATTAT     10140

ATAGAAAACT AGAACTTAGT ACGTATCATT TGTGCGTTTC AATGAGTTCT AGTTTTTTTA     10200

TATGTTATAT TAAACTTATA ACTTTATGGG AGTGGGACAG AAATGATAAA GAGCCACTAA     10260

TGATTTATTA TGTAGTGGTT CTTAAACATT AGCCACAGCT AATGTGTACT TAAAAATAGG     10320

AATACATGAG TAAAACTCAT GCATAAGAAA TACTAATTTC TATAGAAAAA GTATTACTTT     10380

ATCGTTGTCC CACCCCAACT TGCACATTAT TGTAAGCTGA CTTTCCGCCA GCTTCTGTGT     10440

TGGGGCCCCG CCAACTTGCA CATTATTGTA AGCTGACTTT TCGTCAgCTT CTGTGTTGGG     10500

GCCCCGCCAA CTTGCACATT ATTGTAAGCT GACTTTTCGT CAGCTTCTGT GTTGGGGCCC     10560

CGCCAACTTG CATTGTCTGT AGAAATTGGG AATCCAATTT CTCTATGTTG GGGCCCACAC     10620

CCCAACTCGC ATTGCCTGTA GAATTTCTTT TCGAAATTCT CTGTGTTGGG GCCCACACCC     10680
```

```
ACTCGCATTG CCTGTAGAAT TTCTTTTCGA AATTCTCTGT GTTGGGGCCC CTGACTAGAG    10800

TTGAAAAAAG CTTGTTGCAA GCGCATTTTC ATTCAGTCAA CTACTAGCAA TATAATATTA    10860

TAGACCCTAG GACATTGATT TATGTCCCAA GCTCCTTTTA AATGATGTAT ATTTTTAGAA    10920

ATTTAATCTA GACATAGTTG GAAATAAATA TAAAACATCG TTGCTTAATT TTGTCATAGA    10980

ACATTTAAAT TAACATCATG AAATTCGTTT TGGCGGTGAA AAAATAATGG ATAATAATGA    11040

AAAAGAAAAA AGTAAAAGTG AACTATTAGT TGTAACAGGT TTATCTGGCG CAGGTAAATC    11100

TTTGGTTATT CAATGTTTAG AAGACATGGG ATATTTTTGT GTAGATAATC TACCACCAGT    11160

GTTATTGCCT AAATTTGTAG AGTTGATGGA ACAAGGAAAT CCATCCTTAA GAAAAGTGGC    11220

AATTGCAATT GATTTAAGAG GTAAGGAACT ATTTAATTCA TTAGTTGCAG TAGTGGATAA    11280

AGTCAAAAGT GAAAGTGACG TCATCATTGA TGTTATGTTT TTAGAAGCAA GTACTGAAAA    11340

ATTAATTTCA AGATATAAGG AAACGCGTCG TGCACATCCT TTGATGGAAC AAGGTAAAAG    11400

ATCGTTAATC AATGCAATTA ATGATGAGCG AGAGCATTTG TCTCAAATTA GAAGTATAGC    11460

TAATTTTGTT ATAGATACTA CAAAGTTATC ACCTAAAGAA TTAAAAGAAC GCATTCGTCG    11520

ATACTATGAA GATGAAGAGT TTGAAACTTT TACAATTAAT GTCACAAGTT TCGGTTTTAA    11580

ACATGGGATT CAGATGGATG CAGATTTAGT ATTTGATGTA CGATTTTTAC CAAATCCATA    11640

TTATGTAGTA GATTTAAGAC CTTTAACAGG ATTAGATAAA GACGTTTATA ATTATGTTAT    11700

GAAATGGAAA GAGACGGAGA TTTTCTTTGA AAAATTAACT GATTTGTTAG ATTTTATGAT    11760

ACCCGGGTAT AAAAAAGAAG GGAAATCTCA ATTAGTAATT GCCATCGGTT GTACGGGTGG    11820

ACAACATCGA TCTGTAGCAT TAGCAGAACG ACTAGGTAAT TATCTAAATG AAGTATTTGA    11880

ATATAATGTT TATGTGCATC ATAGGGACGC ACATATTGAA AGTGGCGAGA AAAAATGAGA    11940

CAAATAAAAG TTGTACTTAT CGGTGGTGGC ACTGGCTTAT CAGTTATGGC TAGGGGATTA    12000

AGAGAATTCC CAATTGATAT TACGGCGATT GTAACAGTTG CTGATAATGG TGGGAGTACA    12060

GGGAAAATCa GAGATGAAAT GGATATACCA GCACCAGGAG ACATCAGAAA TGTGATTGCA    12120

GCTTTAAGTG ATTCTGAGTC AGTTTTAAGC CAACTTTTTC AGTATCGCTT TGAAGAAAAT    12180

CAAATTAGCG GTCACTCATT AGGTAATTTA TTAATCGCAG GTATGACTAA TATTACGAAT    12240

GATTTCGGAC ATGCCATTAA AGCATTAAGT AAAATTTTAA ATATTAAAGG TAGAGTCATT    12300

CCATCTACAA ATACAAGTGT GCAATTAAAT GCTGTTATGG AAGATGGAGA AATTGTTTTT    12360

GGAGAAACAA ATATTCCTAA AAAACATAAA AAAATTGATC GTGTGTTTTT AGAACCTAAC    12420

GATGTGCAAC CAATGGAAGA AGCAATCGAT GCTTTAAGGG AAGCAGATTT AATCGTTCTT    12480
```

```
GCGTTAATTC ATTCTGATGC GCCTAAGCTA TATGTTTCTA ATGTGATGAC GCAACCTGGG        12600

GAAACAGATG GTTATAGCGT GAAAGATyAT ATCGATGCGA TTCATAGACA AGCTGGACAA        12660

CCGTTTATTG ATTATGTCAT TTGTAGTACA CAAACTTTCA ATGCTCAAGT TTTGAAAAAA        12720

TATGAAGAAA AACATTCTAA ACCAGTTGAA GTTAATAAGG CTGAACTTGA AAAAGAAAGC        12780

ATAAATGTAA AAACATCTTC AAATTTAGTT GAAATTTCTG AAAATCATTT AGTAAGACAT        12840

AATACTAAAG TGTTATCGAC AATGATTTAT GACATAGCTT TAGAATTAAT TAGTACTATT        12900

CCTTTCGTAC CAAGTGATAA ACGTnAATAA TATAGAACGT AATCATATTA TGATATGATA        12960

ATAGAGCTGT GAAAAAAATG AAnATAGACA GTGGTTCTAA GGTGAATCAT GTTTTAAATA        13020

AGAAAGGAAT GACTGTACGA TGAGCTTTGC ATCAGAAATG AAAAATGAAT TAACTAGAAT        13080

AGACGT                                                                   13086
```

(2) INFORMATION FOR SEQ ID NO: 17:

   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 1350 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: double
       (D) TOPOLOGY: linear

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
CATTAGTCAT GAAAATAGCC GACAACTTCA TCTGTGAAAT CACCGGCCTT TTATTTTAGC          60

TAACTTTATT TCTGATTTTA CGATTTTAAT TGATCATACA GAGAAAGTGA TCTTTTTACA         120

ATTTCTAAAA ACTCATGATC TATATTGGAC ATTTGATGAA AATAAGACAA AATGTTTTCT         180

GTTAGCTTCT CTTGTTTTGG GAATGAATCA TCTTCTTTAA TCCAAATCGC TAATTCGCCT         240

AATGGTGTTT TATCATCTTT AAATGTTTGT ATATATTCGT AAAAGCTCAT AGTATTCCTT         300

CTCTCAATTT ACTTATATAA ATCCTACCAC GAAAGCTTTC AAGAAAACAC AATTAAATGT         360

CTATTTAGTG AACTTTTTAA GGTTGTGCAC TCTTTTAATG TCTGCCAATT AGGTCAATTA         420

ATCATCACAA TGTACAATTA ACTCTATTTT CAGTTCATAT ACTCACACAC CGTTTTTGAA         480

CAACACATTA ACTTCTCATT TAGATAAAAC GCAAAAAGC CTGGCACCAA TACAATAGAT          540

GCCAGACTAA GAGTCTACTA TATAAATTTA TTTAGCGTAT GGTTTTACTT CGATTGCACC         600

TTCATTTTCA TCATGAACAC CATGCTTATA ATAATCAATA TATTGTGGCT CTAAAGGCTT         660

TCTGCCACGT ATAATGTCTG CTGCTTTTTC AGCTAACATT AAAACAGGTG CGTGTATATT        720

GCCATTTGTC GTACGTGGCA TAGCTGATGC ATCAACTACA CGTAAATTTT CCATACCGTG        780
```

```
ACTACAAGAT GGGTGTAATG CTGTTTCACC ATCTCTACGA ACCCAATCAA GAATTTCTTC        900

GTCTGTTTGC ACTTCTGGTC CTGGTGAAAT TTCTCCACCA TTGAATGGAT CCATTGCTTT        960

TTGAGATAAG ATATTTCTTG CTACACGAAT TGCTTCTACC CATTCTTTTT TATCTTCTTC       1020

TGTTGATAAA TAATTAAAGC GGATACTTGG TTTTTCGAAT GGATCTTTAG ATTTGATTTT       1080

CAAGCTACCA CGAGAGTTTG AATACATTGG TCCTACGTGA ACTTGATAAC CATGTGCGAC       1140

CGCTGCCTTT TGACCATCAT ATCTTACAGC TATTGGTAAG AAATGGAACA TTAAGTTAGG       1200

ATAAtCAACT TCGTTATTTG AACGTACAAA TCCGCCACCT TCAAAATGGT TAGATGCTGC       1260

TGCACCTGTA CGTGTGAAAA TCCATTGTAA ACCAATAAAT GGcATGCGCT TGAtATCTAA       1320

GCTTGGCtGt AATGATACAG GTTCCTTACA                                       1350
```

(2) INFORMATION FOR SEQ ID NO: 18:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1376 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
TAATGCTATT GGCAACACCA TATATGAAAn CTCCAAACGA TCCTAAACCG ACTATAGATT         60

CACCAAATTT nACAATCCAT GAATAAAGTA GTGGCCATAA GAATAACAAT ATGACAACTA        120

AAAATGTACA GTAAAATGCA GTCATAATTG GAACTAGACG TTTACCACTA AAAAATGATA        180

ATGCTAATGG TAATTCTGTT TCACTAAACT TATTGTATGC ATAAGCTGCT ATTAAACCTA        240

TTACAATACC AACAAAGACA TTGCCATTAT TCATCTTTTC AAAAGCTGAA TTTATTTCCG        300

ArGCTTTCAT TCCTAATAAA GGCGCTAATT TCATTGGTGA TAATACAACT GTAACTAAAA        360

AATATCCTAA CGTrGCTGCA rGCGsGACTG CACCATCATT TTTCTTTGCC ATTCCTATAG        420

CTACACCAAT TGCAAATAAA ATACCTAATT GCTCTAAAAT CGTAGTACCT ACCGTAGTAA        480

AGAACATTGC GATTTTCGGC GTCGCATGAA GTGCATTTAA CGTATTACCA ATTCCGGCAA        540

TAATTGCTGC AGCCGGTAAA ATGGCAACTG GTAACATTAA CGAACGCCCT AAATTTTGGA        600

AAAATTTATA CATTGAATGT CATCCTTCTT AAAATAATGT AGAAATATAA AGATTACTAA        660

TGTAACTAGA ATAACTACTT CGATACTCCG TTATAGTCAC CTAGGCTTAC TAACCAGCTA        720

TATTTCTACC TCAAGTTATT TTATAAACTT TTTACAATTT CATGCAATTC TTGTTGTAAC        780

TTTGCTGTTC GTGTTTCAAT CTCTTTTGTA ATATAATCGA TACGCTCGTT TCGTTTTAAA        840
```

```
AAAGACCGTG AATCTTAGTA GGACCAACAT AAGCAACAGG TAATATTGGT GACTTACTTA    960

ACATTGCAAT TGTTGAAGCA CCaCGTTTCA AAGGTGCACC TTCTTGCGAT GTGCGAGAAC   1020

CTGTTGGGAA GATACCAACT GTCTTATTAT CTTTCAACAA ATTGATTGGG CGTTTTAAAG   1080

TACTAGGTCC TGGATTTTCA CGATCTACAG GAAATGCATT TAAAGACGTT AAAAATTTAC   1140

CAATCCATTT ATTTTTGAAT AATTCTTTTT TAGCCATATA ATGAATTTGA TTAGGATATA   1200

ATGCCATACC TAGCATAATG ACTTCGTTAT AACTTTCATG CGTACAAGTT ACGACATATT   1260

TACTATCCTT AGGAATATTA TCTTTACCGA TTACGTATAA TGATTTTGAC ATTTTAACTA   1320

AAATGAAATT CAAAATCTTA CTAATCACTG AATACATTGT GCCACCTACT TAACTT       1376
```

(2) INFORMATION FOR SEQ ID NO: 19:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7363 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

```
TTGTCATACC AATATTTTGT AAAATATGGA ACACAAGTAA AGTGACGAAA CCAACGATAA     60

AGATTTTGTT AAATTGATCT TCAATTTTCG CAGCTAATCT TATTAGATGG AAGATTAAAA    120

ATAAAAATAT TAAGATCAAT ATGACAGAAC CGATAAAGCC AAGTTCCTCT CCAATCACTG    180

AAAAGATAAA GTCAGTATGA TTTTCAGGTA TATAAACTTC ACCGTGATTG TATCCTTTAC    240

CTAGTAACTG TCCAGAACCG ATAGCTTTAA GTGATTCAGT TAAATGaTAG CCATCACCAC    300

TACTATATGT ATAGGGGTCA AGCCATGAAT TGATTCGTCC CATTTGATAC AGTTGGaCAC    360

CTAAJAAATT TTCAATTAAT GCGGGTGCAT ATAGaATACC TAAAATGACT GTCATTGCAC    420

CAACaATACC TGTAATAAAG ATAGGTGCTA AGATACGCCA TGTTATACCA CTTACTAACA    480

TCACACCTGC AATAATAGCA GCTAATACTA ATGTAGTTCC TAGGTCATTT TGCAGTAATA    540

TTAAAATACT TGGTACTAAC GAGACACCAA TAATTTTGAA AAATAATAAC AAATCACTTT    600

GGAATGATTT ATTGAATGTG AATTGATTAT GTCTAGAAAC GACACGCGCT AATGCTAAAA    660

TTAAAATAAT TTTCATGAAT TCAGATGGCT GAATACTGAT AGGGCCAAAC GTGTACCAAC    720

TTTTGGCACC ATTGATAATA GGTGTAATAG GTGACTCAGG AATAACGAGC AAGCCTATTA    780

ATAATAGACA GATTAAGAAA TACAATAAAT ATGTATAATG TTTAATCTTT TTAGGTGAAA    840

TAAACATGAT GATACCTGCA AAAATTGCAC CTAAAATGTA ATAAAAAATT TGTCTGATAC    900
```

TTGCTAAAAC AGCTATAGTG GCTACTAATA CCCAGTCTAC TTTGCGAAnC aATGCTTATC   1020

CGGCTGTTGA CGAGATGAAT AATTCATTGC AAACTCCTTT TATACTCACT AATGTTTATA   1080

TCAATTTTAC ATGACTTTTT AAAAATTAGC TAGAATATCA CAGTGATATC AGCTATAGAT   1140

TTCAATTTGA ATTAGGAATA AAATAGAAGG GAATATTGTT CTGATTATAA ATGAATCAAC   1200

ATAGATACAG ACACATAAGT CCTCGTTTTT AAAATGCAAA ATAGCATTAA AATGTGATAC   1260

TATTAAGATT CAAAGATGCG AATAAATCAA TTAACAATAG GACyAAATCA ATATTAATTT   1320

ATATTAAGGT AGCAAACCCT GATATATCAT TGGAGGAAAA CGAAATGACA AAAGAAAATA   1380

TTTGTATCGT TTTTGGAGGG AAAAGTGCAG AACACGAAGT ATCGATTCTG ACAGCACAAA   1440

ATGTATTAAA TGCAATAGAT AAAGACAAAT ATCATGTTGA TATCATTTAT ATTACCAATG   1500

ATGGTGATTG GAGAAAGCAA AATAATATTA CAGCTGAAAT TAAATCTACT GATGAGCTTC   1560

ATTTAGAAAA TGGAGAGGCG CTTGAGATTT CACAGCTATT GAAAGAAAGT AGTTCAGGAC   1620

AACCATACGA TGCAGTATTC CCATTATTAC ATGGTCCTAA TGGTGAAGAT GGCACGATTC   1680

AAGGGCTTTT TGAAGTTTTG GATGTACCAT ATGTAGGAAA TGGTGTATTG TCAGCTGCAA   1740

GTTCTATGGA CAAACTTGTA ATGAAACAAT TATTTGAACA TCGAGGGTTA CCACAGTTAC   1800

CTTATATTAG TTTCTTACGT TCTGAATATG AAAAATATGA ACATAACATT TTAAAATTAG   1860

TAAATGATAA ATTAAATTAC CCAGTCTTTG TTAAACCTGC TAACTTAGGG TCAAGTGTAG   1920

GTATCAGTAA ATGTAATAAT GAAGCGGAAC TTAAAGAAGG TATTAAAGAA GCATTCCAAT   1980

TTGACCGTAA GCTTGTTATA GAACAAGGCG TTAACGCACG TGAAATTGAA GTAGCAGTTT   2040

TAGGAAATGA CTATCCTGAA GCGACATGGC CAGGTGAAGT CGTAAAAGAT GTCGCGTTTT   2100

ACGATTACAA ATCAAAATAT AAAGATGGTA AGGTTCAATT ACAAATTCCA GCTGACTTAG   2160

ACGAAGATGT TCAATTAACG CTTAGAAATA TGGCATTAGA GGCATTCAAA GCGACAGATT   2220

GTTCTGGTTT AGTCCGTGCT GATTTCTTTG TAACAGAAGA CAACCAAATA TATATTAATG   2280

AAACAAATGC AATGCCTGGA TTTACGGCTT TCAGTATGTA TCCAAAGTTA TGGGAAAATA   2340

TGGGCTTATC TTATCCAGAA TTGATTACAA AACTTATCGA GCTTGCTAAA GAACGTCACC   2400

AGGATAAACA GAAAAATAAA TACAAAATTG ACTAACTGAG GTTGTTATTA TGATTAATGT   2460

TACATTAAAG CAAATTCAAT CATGGATTCC TTGTGAAATT GAAGATCAAT TTTTAAATCA   2520

AGAGATAAAT GGAGTCACAA TTGATTCACG AGCAATTTCT AAAAATATGT TATTTATACC   2580

ATTTAAAGGT GAAAATGTTG ACGGTCATCG CTTTGTCTCT AAAGCATTAC AAGATGGTGC   2640

TGGGGCTGCT TTTTATCAAA GAGGGACACC TATAGATGAA AATGTAAGCG GGCCTATTAT   2700

268

```
AAACCCTAAA GTAATTGCCG TCACAGGGTC TAATGGTAAA ACAACGACTA AAGATATGAT    2820

TGAAAGTGTA TTGCATACCG AATTTAAAGT TAAGAAAACG CAAGGTAATT ACAATAATGA    2880

AATTGGTTTA CCTTTAACTA TTTTGGAATT AGATAATGAT ACTGAAATAT CAATATTGGA    2940

GATGGGGATG TCAGGTTTCC ATGAAATTGA ATTTCTGTCA AACCTCGCTC AACCAGATAT    3000

TGCAGTTATA ACTAATATTG GTGAGTCACA TATGCAAGAT TTAGGTTCGC GCGAGGGGAT    3060

TGCTAAAGCT AAATCTGAAA TTACAATAGG TCTAAAAGAT AATGGTACGT TTATATATGA    3120

TGGCGATGAA CCATTATTGA AACCACATGT TAAAGAAGTT GAAAATGCAA AATGTATTAG    3180

TATTGGTGTT GCTACTGATA ATGCATTAGT TTGTTCTGTT GATGATAGAG ATACTACAGG    3240

TATTTCATTT ACGATTAATA ATAAAGAACA TTACGATCTG CCAATATTAG GAAAGCATAA    3300

TATGAAAAAT GCGACGATTG CCATTGCGGT TGGTCATGAA TTAGGTTTGA CATATAACAC    3360

AATCTATCAA AATTTAAAAA ATGTCAGCTT AACTGGTATG CGTATGGAAC AACATACATT    3420

AGAAAATGAT ATTACTGTGA TAAATGATGC CTATAATGCA AGTCCTACAA GTATGAGAGC    3480

AGCTATTGAT ACACTGAGTA CTTTGACAGG GCGTCGCATT CTAATTTTAG GAGATGTTTT    3540

AGAATTAGGT GAAAATAGCA AAGAAATGCA TATCGGTGTA GGTAATTATT TAGAAGAAAA    3600

GCATATAGAT GTGTTGTATA CGTTTGGTAA TGAAGCGAAG TATATTTATG ATTCGGGCCA    3660

GCAACATGTC GAAAAAGCAC AACACTTCAA TTCTAAAGAC GATATGATAG AAGTTTTAAT    3720

AAACGATTTA AAAGCGCATG ACCGTGTATT AGTTAAAGGA TCACGTGGTA TGAAATTAGA    3780

AGAAGTGGTA AATGCTTTAA TTTCATAGAG ATTAGTCGAG GGACCTTTTA CTTATAAAAA    3840

TGATTTGAAT TAATACTAAA AGATTACAAA GAAGAGGTGG TTTTGTGTGT AAATACAAAA    3900

TTGCCTTTTT CTTTTTATGT TAAATCTATA AATTTGAAAC TAAATCAAGG TTAATTCTAT    3960

GTACACACTT TATATAGGAA GTAGTTTGAA TGTTTATATA ATGTTTTACA AAAAGATGTA    4020

GTATTATAAT GTCTAATTTC ACATGTGTTT CAGTAAAATT TGTTGTGGAA TGTTAACGAT    4080

ATACGTATTT TATAAAAaAT TTTTTATAAT GATTATTCGA ATGATGCGTA ACGCTTACAT    4140

CTTATCTAAT GCTAGCTTTT TGACAAAAAT ATGACAATCA ATTAATGTGA TTCTAATAAA    4200

TATTCGCAAA TTGCTTTATT GCGATTAAAT TTTTTTGGTG GTACTATATA GAAGTTGATG    4260

AAATATTAAT GAACTTATAT GCAAAGTAT ATTGAGAAAT AAACAGGTAA AAAGGAGAAT    4320

TATTTTGCAA AATTTTAAAG AACTAGGGAT TTCGGATAAT ACGGTTCAGT CACTTGAATC    4380

AATGGGATTT AAAGAGCCGA CACCTATCCA AAAAGACAGT ATCCCTTATG CGTTACAAGG    4440

AATTGATATC CTTGGGCAAG CTCAAACCGG TACAGGTAAA ACAGGAGCAT TCGGTATTCC    4500
```

269

```
AGAATTGGCA ATGCAGGTAG CTGAACAATT AAGAGAATTT AGCCGTGGAC AAGGTGTCCA          4620

AGTTGTTACT GTATTCGGTG GTATGCCTAT CGAACGCCAA ATTAAAGCCT TGAAAAAAGG          4680

CCCACAAATC GTAGTCGGAA CACCTGGGCG TGTTATCGAC CATTTAAATC GTCGCACATT          4740

AAAAACGGAC GGAATTCATA CTTTGATTTT AGATGAAGCT GATGAAATGA TGAATATGGG          4800

ATTCATCGAT GATATGAGAT TTATTATGGA TAAAATTCCA GCAGTACAAC GTCAAACAAT          4860

GTTGTTCTCA GCTACAATGC CTAAAGCAAT CCAAGCTTTA GTACAACAAT TTATGAAATC          4920

ACCAAAAATC ATTAAGACAA TGAATAATGA AATGTCTGAT CCACAAATCG AAGAATTCTA          4980

TACAATTGTT AAAGAATTAG AGAAATTTGA TACATTTACA AATTTCCTAG ATGTTCATCA          5040

ACCTGAATTA GCAATCGTAT TCGGACGTAC AAAACGTCGT GTTGATGAAT TAACAAGTGC          5100

TTTGATTTCT AAAGGATATA AAGCTGAAGG TTTACATGGT GATATTACAC AAGCGAAACg          5160

TTtAGAAGTA TTanAGAAAT TTAAAAATGA CCAAATTAAT ATTTTAGTCG CTACTGATGT          5220

AGCAGCaAGA GGACTAGATA TTTCTGGTGT GAGTCATGTT TATAACTTTG ATATACCTCA          5280

AGATACTGAA AGCTATACAC ACCGTATTGG TCGTACGGGT CGTGCTGGTA AAGAAGGTAT          5340

CGCTGTAACG TTTGTTAATC CAATCGAAAT GGATTATATC AGACAAATTG AAGATGCAAA          5400

CGGTAGAAAA ATGAGTGCAy TcGTCCACCA CATCGTAAAG AAGTACTTCA AGCACGTGAA          5460

GATGACATCA AAGAAAAAGT TGAAAACTGG ATGTCTAAAG AGTCAGAATC ACGCTTGAAA          5520

CGCATTTCTA CAGAGTTGTT AAATGAATAT AACGATGTTG ATTTAGTTGC TGCACTTTTA          5580

CAAGAGTTAG TAGAAGCAAA CGATGAAGTT GAAGTTCAAT TAACTTTTGA AAAACCATTA          5640

TCTCGCAAAG GCCGTAACGG TAAACCAAGT GGTTCTCGTA ACAGAAATAG TAAGCGTGGT          5700

AATCCTAAAT TTGACAGTAA GAGTAAACGT TCAAAAGGAT ACTCAAGTAA GAAGAAAAGT          5760

ACAAAAAAAT TCGACCGTAA AGAGAAGAGC AGCGGTGGAA GCAGACCTAT GAAAGGTCGC          5820

ACATTTGCTG ACCATCAAAA ATAATTTATA GATTAAGAGC TTAAAGATGT AATGTCTTGA          5880

GCTCTTTTTT GTTTTCAATA ATTGATTCTC TGTAGATATC aAAGTaCTAA CGTTTTAAAG          5940

GTTAAATATT TAATTGGATT GAGATCTGTA TGCGGTTATA TCaTTCTGTG TAAATATGGT          6000

TCTCCACCAA ATGTGGTGAG TATATAATTT AAAGAACTAT TTTTAAATTA AGAATAATCG          6060

AACATAAATA AACTTTATGA AATTTCAGTA TCATGTTCTT ATAAAAAACA ATAGGGCTTT          6120

TTGCtGACGC TAGTGCGCGA TAAATAATAA GTTGAATATA AAAAGATCA CTGCCAATCA          6180

TTCGTTTAAT GGCAGCGATC TTTTTTATTT AATTATTTCT CTTTCCACTG CAACATTTGA          6240

TAACCAATGC GTGGATGTGT TTTAATAATA TCTTTTGCGT CCTCATGACA TTGTGAAAGT          6300
```

```
CCATATATTC GTTTTAATAT CATCTCATAA GTGAGTACTT TTCCTTTATG ATTTGACAAT      6420

AGTTCTAACA AGCTAAATTC ATTTGGCGTC AAATGTACCT CCTGATTATT AATAACAACA      6480

GATTTGGAGC CAAAGTCGAT GCTTAGCAAA CCGTTAGTAA ATACAATGTT AGTTTCTTGA      6540

TGTGACTTAG CGATTCTCTC GATGACTCGT ATTCGTGCCC GAAGCTCATC AACATTAAAA      6600

GGTTTAGTCA TATAGTCATT CGCACCGTTA TCTAAAGCTT GAATAATTGT TTGTTCTTCT      6660

TGTCTTGCAC TTATTACAAT GATAGGAATG TCAGTATGTT GCCTGATTTC TGAAATCAAA      6720

CATAATCCAT CTTTATCTGG TAAACCTAAA TCTAATAAAA TGACATCTGG TTTATCAATT      6780

TGAATTTTAA AGTGTGCTTG TGTGGCATTG TCGGCTGTAG TTACATTGTA ATAATCTAAA      6840

GTTAATGCAA CATCAAGTAA ATGTGTGATT GCGTGATCAT CTTCAATTAT CAATATTTTA      6900

GATTGCATTA TACGTCTCCT TCGTTAAAGT CTGTATATAT ATTGAAATAG AATATACTGC      6960

CGTGTGGTTG GTTCGGTTTA TATTGTAAGT TTGATTGATG TTTGTGTAGG ATAGTCTGTA      7020

CTAAATATAA GCCTAGTCCC ATGCTTTCTT TTTGGTTATC TTTAAAATAT TTATTTGATC      7080

CTGTGTAAAA AGGCTCGAAT ATCTTTTGTt GTTCTTCTAA ACTAATTCCA GGTCCTTCGT      7140

CTATAACGGC AAATTCGATT TGTTCATAGC TAGCATAACG AATAGATAAA TTGATTTTGG      7200

TGTCAGTAGA AGTGTGTTTA ACTGCATTTT CAATCAAATT GAAtAAAgCT TGTAAAATCA      7260

ACTTACTGTC AATGTGTATA AACtGTAAAT TTACTGAGGA TGATACAGTT ATACGCTTTT      7320

TTAAATGGCG ACGTTCTAAA ATACATATCG ATTTCTTATA CTA                        7363
```

(2) INFORMATION FOR SEQ ID NO: 20:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10470 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```
TTAACAATCG ATAACCACAA TACTTCTATT GTAATTGTTT AACGATTTCn CGATTAAAAT        60

CATCTAAATC GTCTGGTACT CGACTTGTTA CAATATTGTT GTCTACAcTa CTGACTCATC       120

AACTACATGT GCGCCTGCAT TTGATAAATC TTTGCGTACA TTTAATACTG CTGTTAACGT       180

ACGACCTTTT AAATCGTCTG TATCTATTAG TATTTGTGGC CCATGACAAA TGGCAAATGT       240

TGGTACATCA TTTTTAGTAA AGTATTTAGC AAATGTGCCA TATCGACCTT CTGTATCTCC       300

ACGTAAATGA TCTGGTGAAA ATCCTCCAGG AATTAATAAT GCATCATAAT CTTCTGGTTT       360
```

ATTTGCAGTA TCTCCAATCA CTACAGTATT AAAGCCTGCA TTCTCTAATG CCTCTTTAGG   480

GCTTGAATAT TCTATATCTT CAAATTCGTT TGCTAGAATA ATTGCTACTT TTTTAGTCAT   540

TGAAAATCAC CTTTCTATAT ATCATTGATA TAATTACTAT AGACAAGTAA ATCAGTGATT   600

AAACATACAA GATATAAAAA ATATTAAGCG ACTGTCGCGA TATCTAACCC TAACACATCT   660

TATGTGGCAT TTACTTAGAT ACTAATTTAA CCTTTTCTTC AAGCTGATCT AACAATCCAA   720

TCCATTCATC TATATCTTCA ACACGTACTT CATCAGGATT TACATGATCG ATATCCTCAA   780

TAAACTTATT TAAACGCGCT TTTATCTGTT CGATTGTTTG CTGTTCATTC ATAAAAAGTT   840

AACTCCTTTT ATTTTGTTTT CTTTTTCATT ATTATCCTAA CAGAAATTGC GTTAAAGCGA   900

TATAATCTTA GCTATATTTA TGACATTCAA ATTATTTTGA CTTTTAAAAA TCCCCTTTTC   960

AATTAACTAA AATTAAGAGA TAATTTGTTA CGAGTGATAA TACGAaGkGG TaTCATACCG   1020

ATATGAACCA AATAGAAAGA AGGAAGTTTA AGACGATGAA TAGCGTCAAA TTGAAGCAAC   1080

CTGTTAGCAT TTACAATGAT CCATGGGAAG TGAAATTTAT ATACATTTAA ATTTCATGAG   1140

ACAATAAACG TTGATTTAAT GCGTTTTTTT GCCTTTTTTA TTTTCCTTAT TTTTTCTGTT   1200

TTACAACAAA ATGGTATCAA AAATGGTATC ATTTGTAGTT ATTTTAGCTT CACATATTAA   1260

AACAACCACA CTCCTAAATT AATAGGTGGT GTGGTTTTGT TGGTTGTGTG GGGATAAAAA   1320

TAACCGCATC AGTTAAGATG CGGTTATCTA GCAAGGGCCA CGTATTTATA AATACGTTTA   1380

GAATCTCTTC GGCAACTTTG CTATAGACAG TCTATGCTGT TACTAAATTA TACCACCACA   1440

CAAACCTACT CCCATTCAGG AACACAGAGC TTTGTCGCTC GTCAGCAACG TCATATGAAT   1500

TCTCAGTTCA TGTTGTGGTG ACACTTTAAA CGGTCTGTGC CAGTAGCGAC CGAGTCATTT   1560

CAAGAATGAC CATTTCACAT TTATATTATA ACACTTGTCG TGCGTAACTG TATAGTTTTT   1620

CAGTTGTATT TAAAGTTAAG TTATCTACTT CGCGCTTTCC TTGCCTTAAT TGTGAAATTA   1680

CATATTGCGC TACGCCAGTT TGTTTGTGAA TTTGGTAACC TGTTATATCA CTTTTGATCA   1740

ATTCAATTAT TTTTAATTTA TAATCACTCA TATTATCTAC GTCCATTCTT TTTATCTAAA   1800

CAATAAAAAT GTGTCTTTCT CCCGATAAAT AATAACAATG GTAGGCTTAA TAAAAACAAT   1860

ATTAAATACA TTTGTTCTGT CATAATTGAA AACCTCCAAA TAATATTATA TTATATAAGT   1920

GTAAGGAGGA GCCATCAGGC TCCAAGCATA ATGTTAATCT TTGTTGTTTG GCTTTCGGTC   1980

TAGGTAGCCG AGATGCCaTT CTCTAAGTTG TTTTAACACT TCTGGAATTA TCAGTACTGC   2040

CAATACTTGA TGTTCTAGAA GTGTTTTTAT TATGTCTAGC ATGAGGCTTT TCACCTCCTT   2100

ACACATAATT TGTAAGTCAT CAACTAACCT ACAAATATAA TTATACTAAA CAAATGTTTA   2160

```
GTTATCTACA  TTTAAATCTT  GAGAGAAATG  TTAAAAAGTT  CTAGTAAAAT  AATAGCACAT    2280

TTTATCTTTA  AATGTAAATA  GAAAGCAGGT  ATGTAACGCA  CCTGCTTAAA  TAGaCATGAC    2340

TATGTCATTC  TAACTGATTT  CTCCCCATAA  GTCACCTAAT  ATCTGATTAG  GTGGGGCAGA    2400

ACCATTCCAT  GTTCTAATAG  GCAAGTAATA  ACGTTGCCCC  TCCCATGTAT  ATCCTACCCA    2460

AACATGACCA  TCTTGTAACA  TCACTTCTGT  ATAATCACAA  TACCCACCAG  GTTGGAACTG    2520

ATAACCCACT  GGACAAGATA  AGAATGGCCC  CACTTTTCTT  ACTGTGATTG  GTTGATTGCC    2580

GTTTGTGAAT  CTAGCACTTT  CTTCCATGTA  GTAAGTACCA  TATTTATTAC  GTTTCCATGC    2640

ACTTGCAACT  GGTTTAACTG  TATTACTTGA  AGCGCTTGAC  TCATTAGAGA  CAGTGGCAAC    2700

CGGTATTTTA  CCATCCATGT  ACGCCCTAAT  CTGCTTGATA  AAGTAGTCTT  TAAGTTGCAA    2760

CCGCTTGTCT  TCTGGCAATA  GACCGCGAGT  TACTGGGTCA  AAACCAGTGT  GTAAAACCGA    2820

ACTTCTATGA  GGGCATGATG  TTGAAGTAAA  TTCATTGTGC  AATCTGATTG  TATTTCTGTT    2880

TGCTGGTAAT  CCCCATTTTT  TCAACAATCT  AGCGCATTCT  TGGAAAGTTG  CCTGTTCATT    2940

TTTTAAGAAT  GTCGCGTTAT  CTGCGCCCAT  TGATTGACAT  ACTTCAATAC  CGTAATAATA    3000

TTTATTACCT  ATTTGATTAG  CGGTATGCCA  ACCTACTTGT  GATTCATCTA  AGGCTTGCCA    3060

AACTGTGTTG  CCTGATACGT  AACTATGCGC  AATGCCCGCT  TCTAATCTTG  ATAAAGGTGC    3120

ATTTACTAAT  CCGTTACGAT  ATGCTTCAGC  AGTCGCCCCT  TTGCTCCCTG  CGTCGTTGTG    3180

TATAACTATA  CCTTTAGGGT  TACTACCACG  CTTAGGTAGG  TCATAACCTT  TAACCACATC    3240

TTTGATGATT  TTAAGTTCTA  CTGCTTTAGG  TTGTGGCTTA  GCTGTTTCTT  TTTTAGGTGC    3300

TTGTGTAGGA  GATTGAACTG  ATCGTGGCGC  TGTCTCACTT  TTAAAATTCG  GACGGATAAA    3360

CCACATAGGG  AAATCATAAG  CATGTTGTCG  TCTTGTAACT  TTTTCCCAAC  CCCAGCCGGG    3420

TTGTTCGATT  CCGTCAGTCC  AGCCACCGCC  TAGCCAATTC  TGCTCATATA  CAATGATGTA    3480

ATCTAAAGTT  GCTTCAATTA  CCCATGCAAC  GTGACCATAT  CCAGCACCGT  AGTTGCTACC    3540

GAATACCACC  ATGTCGCCAG  GTTGTGCTAA  GAAGTCCGGT  GTATTTTGGT  ATACAGTAGC    3600

TAATCCGTCG  AAGTTGTTAG  CGAACGGAAT  ATCTTTTGCA  CCTAAACCTT  TTAGAAGTAA    3660

TCCAAACAAA  ACTTTCCAAC  CAGCATTGGC  ATAATCAAAG  CATTGAAATC  CATACCATAA    3720

GTCCACATTG  AATTGTTTTC  CCTCAGAAGT  TTTCAACCAC  TCTATAAACT  CATTTTTAGT    3780

TAATTTTGCT  TGCATTGTCG  CCACCTCCAT  GATGATACTC  ATTCACATCA  AAGCCAACAT    3840

CGTTAGAGGC  GTCTGTGAAA  GGTTGTGATG  TATCATATTC  TTTTGGTGcT  TTCGCGCTTA    3900

ATTCCGGCGT  TAAACTACTG  TCTTGTGATG  ATTTCCACGT  AACTTGTTGT  TCTTCTTTTT    3960
```

```
TTGGGTCAGT AATAACGCCA ATACCTGTAA GTAACGTGAG GATAGCGCCT ATAATTGCGC    4080

TAGCTTGATT TAATTGAGTA GATAAATCTA ATCCGAATAA ATCCGTGACT TGCTTGATAA    4140

ATAGCAACAA TGCTCCAACT AAACCAGTTA GTACTGCTTT GTTTTTGAAT CTCAATTTCC    4200

AGTTAATATC CATTTGTTTG CTCCTTTTAT CCAAAATAAA AAAACGACTA AAAATTAGTC    4260

GTTTAAAATT ATTCAATGGT CAATGTCGGA GATCCTGAAT AAACATCACT TATAGTGACG    4320

TACAACATCC CTGAAGGATT ACTAAAGTTG ATATTTTTAC TTGCAACTCC GCTATTGACT    4380

CCTGATATTC CTAAATCACT TGACCCTAAA TTAGTTTGCG AAATCCTCAT TATACCGCTA    4440

CGTACATTTT CTATTGTCAC CTGATAACTT TTATTGGGTT CAACTCCATT TATTGTCCAT    4500

TTTGCTGTTG ATTCTTCTAT GCTATCCGGA TATTTATTTT TAGGTAAGGG TTTTATTACA    4560

AAAGATGAAG GCTTTTTCCA TACTTGGATA TTTCCAGCAT ATACTTTTGT ATATTCTTCA    4620

CCTTCGTAAA TAAACTTCTT TACATTTTTA AAATTACCTT CCATAAAAAT CACCCTTTAA    4680

TTAAATATAA CGTATTCGGG TCTTTTTGAT ATATATAGTT ATATTCATTT TCTGTTCCTG    4740

TCCAAATTTT AACCGTCGGT TGAGATGCGC TTTTTAGTTG ATATAAATTA TCCGCTTGTT    4800

GTTTAGTAAA AGCTTGAGAT GACAAAACAT ACCGCTCGTC ATGATTATGA TTTTTTGGAG    4860

CATATAAATC ATTTAGTGTT TGTTTGAATT CCTCAAAATC TTCTGTATTA ACTTTTGAGC    4920

CAATCTGTTG CAATACACTT TCTGAAATAG AGTTGTTTTG TATTGCTTCT GCTAATTCTC    4980

TTAATGTGTT CATAGATTCA GGCGCGCTAT CAACTAGTTC AGCAATTTTT GTATCCGTAT    5040

ACGTTTTAGA GTCGTTGAGA GTTGTATCTT TGATTTTTTC AACTTCTTGC AATTTATTTT    5100

CTAACCCTTC AACATTTGCG ATATTGATTT TGTCCAATAA CTCAGGTTCT GCTTTGATAT    5160

CTGTATCTTT ACCATCAATT TGCCACATTT TAGTGTCAGG ATTGATTGAT ACTACAGTAC    5220

CGTTTTTACC GGGTGCGCCT TGTTCTCCTT TTTTACCTGC TTCACCTTTT GCTCCAGGTT    5280

GTCCCGGTTC ACCTTTATCA CCTTTCGCAC CTTTAAATCT ACTTTCATTC TTTTCGATGT    5340

AAGAAATGAC ATCTTTATCT ATTTTCTCTT TAAAGTCTTT GCTCAATAAA TCTGTCGCGT    5400

TATCTTTTAA AATTCTCGTA ATAGCATCAT CTACCAATTT AACATCGATT TCTTTTGCTA    5460

CAGCAGATTC AATACCACTA TCAACGATAT TGAAAGAAAA GTTTGCGACA TGTATTTTTT    5520

CTTCTTCTTT CTCTAAAAAC AGCTTACAGC GAACATAACC AGCGTGTTTG ATAACCTTTT    5580

TAGGTATCTT GTAGGTAAGG AAACCTTTTA CAACATCGTC GATAATAAGG GGCTCATTTT    5640

TGAATATAGA GCCATCTTCC ATAAACAAAT GTAATCTAGG TGTTAAGCCA TGTGCTTTTA    5700

GATCGATACG ACCTTGTTTG TCATTGATAC CTATTCTTAT AGATGCTGTA TTTTCATCTT    5760
```

```
CAACATCTTT TATTTTGTAC ATTTACACAC CTCTTTATTT ATATTTATCC CTTGTGAAGT    5880
AGATACCTTT TAAGCCGATT TGTTTATATA ACTTAGCGAT TGTACTTGCT TGATGTTGGC    5940
ACCACTCTAT AGCAGTAGCG TATTGGTGGG TAGCTGGATT CTTAGGATTC CATCTAATTC    6000
GGTACAATGT GTTTTGACCT TTATTGATGT AATCCTTTCT TACGAAGCTA GCACCGCCCA    6060
TGATTGCTTT TGCTGGAGAT GTCCAACCTT TATTCCTTGC AAACGTCATT GCGTAGTTAG    6120
GATTGTTGTC GTAAGCGCCA ATGCCGAAGT AGTTGTATAC TCCATCTTTT CCGTTAGCGA    6180
AGTTACTTGT TCCATATCCA CTTTCTAAGA AAGCATGCGC GATTAAATAA ATTTCATTAA    6240
TGTTGTGCTT TTTACAAGCT TCTGCGAACG CTTTACCTTG ATTATTCAAT GTTCCCTTAC    6300
CTTTAAGTAT CTTATTAAGT GCGCTAACTG AAACACCTTG ATACTTGCCT AAATTAAGCA    6360
TTTGGTAGCA TTGTGTGTTA CTTTCCCATA TACGCTTTAC ATTCATTGCT GAACTCGTTT    6420
GTGCTCGTGT AGCGTTAscC AACCCCAAGC ATTAGATTTT TTCGGGTTAC CTCTTGCCAT    6480
TTGTTTATCC AGTGCTTGTT TGAATGTATA AGGACTCGTT TCTGTTATGA TCTGCGGTTG    6540
TTTAGATGCC GAACCATTGT TGGCTGTTGG TGACGAGTCT CTTACATTAG CTATATCAGC    6600
GTTTTTATTA TCTACCATAA CTTTTATTCT AGATTTTGTT ACTGTTGGCT TAGTTATAGA    6660
ATTTAATAAT TTTTCTCTGT TTTTAAATAT ATTAAGTAAT GCCTTTTCTA ATGCTTCGTA    6720
TTTATCTTTA GGAGGAACAC CGTTGTCAAT CATATTCCAA TTAACATGTT CCAACATTGA    6780
ACGCCAAATG CTGTCGTCTA CTTTTAAATT TTCAATACTT AGAGGTATCT CATATTTGGC    6840
CATCATATCT ACAGCTACAA CCATTGCGTG AATCTCATTA AAAATAAATT CATTTTTACT    6900
CGCACTATAA TCTTCACATA CGTCTATAAC TATATAATCA GGTTCATTAG GAACTTCAAA    6960
TACAGCTCTT CTAGGTGCCC AAATATTATG TCTATCAACA TAAAAGTGGG GATATTCTAC    7020
ATCCTGTTTG TATTTCTTCC TACTGTTATA TAAACTTTCT ACCGAGCTCA TCGTTTGTGC    7080
GTTTCTAATC ATTATTCCTT TAGGTTTTTC GAGTCGTCGA TTACCTTCTA CTATAAAGTG    7140
ATAAATATAT TCTGGATAAT TAACCTCTTG GCTAGAAATA GTGTACTTTA TAGTTGTTAC    7200
ATCTTTCCAA ATTGGAACTT TTTTATTATT TTTTTCGTTA TCATCACTAT CATCTTCTGG    7260
TTTAGGTGCC GGTGTAGTTT TGTCTGGATG ATATGGTGGT CTAACAAAAT ATTTAACCCC    7320
TCCACCTGGT CCATCATGAT AAGAGTGTTT AATTTTATAA GGTGGACTTC CTGTTGCGTT    7380
ATTTGTATAC CAGTTTTGAT CTACGCCATA CCAATAGTCT TTTGTGCATG GTCCCACTAC    7440
AATGTTTACA TGTCCTGCCC AACCACCAGT CCAAACACCC CAGTCGCCTG GTTGTGGTAC    7500
AAAATCTTTT GTATTTCTAA TTATCTTGAA ATCTCTACCT CTATAATTGG ATTTTTGAGC    7560
```

```
TAAATCCCAG CATTGTGCTC CCATTCCAGA ACCAGGTACA TCAATAGCTA TTTTGTTTTT    7680

AGCGATATAT AACGCCCATT CAACCACTTC ACTAGCTGTG GGCTTTCTAT TTTTCGGATT    7740

AGGTAATCCC ATGTATGCAC CTCATTTCAA TCAAAATAAA AAGCCAGTGC CGAAGCACTG    7800

ACTCTTAACT GTTATTTACA TTTACCAAAC CAGAAGCACG CCCAGAAGCT ATATCCTAAA    7860

ATCCCTTTAA GCATGGTAAT CACCTCCTTT AAATACCAAA AACAGTTCTT AGTAAAGCTA    7920

TGACAATCGT ACTGAAGATA GTCCCTATCA AACCTAGAAT CCACATTTTT ATGTCTCTAA    7980

TATTCTTGGC ATTCTTTTCT TTATTCTTTT CATCTTCTAC CTTGTCGCGC TTTAATTCTT    8040

CAAAATTTCT ATCTAATTTG TCATAAATCT TTTCTTGCGC TCTAAGACTA TCTTCTATTC    8100

TGTCGAATTT TTCAAACATA GTCTTATCAT TTTCTTCTAA TCGCGTTAAA CGCCAATCTT    8160

GTTCATGTCG TTTGGTAAAT CCAAACATTA TGCCACCCAC TTTATTCAAA TTAAAAAGCC    8220

ACAAGCATTA CACCTGTGAC TTTTCATCTT TTGTTTCTGG ATATTTTTCT CCAGTGATTA    8280

AAGCGTATTC TTCTTTATCG ATTAAACCCT TGTCTACGTA CCACTTAATT TGCTCGTTTT    8340

TATAGTAACC CCAAACATAA AAAGTTTTAA TGTCTTTAAA AGTTGGATAA ATCATCTTCA    8400

TTATTTAAAC GTCCCCCTCA GTACTTGTTT TGTTAGTTTT CAGTTCAGTC AACTGTTGTG    8460

TTAACATAGC GTTTTGTTGA GCTAATTCCA TTGTTAATAC GTTTACTTGT GCCACCTGCA    8520

TTTGCATACT CGCAACCATT CCGCGAAGTT CCTCATCACT TAAATCTGAC GCACTTTGTT    8580

GGTTTGATGC ATTCGGTACG TCTTCTTTTT CGAAATTGCT ATTGTATTTA ATTTCGCCGT    8640

TAGTGAAAAC AAACTTTCTA GGTTCGAACT CTTCTTTAAA TTTAATAGGC ACATTGTTAT    8700

CATCTACATC TAAACTATTG CGTAAACCGC CAGTATTAAC GAATCCGATA ACTTCGTTTT    8760

TATCGTTTAC TGTGATTTTC ATTATTTCCA CCCCATAATT TTAGTTATAG TAACTTTGTT    8820

GGCATTCGCT CCAGAACCTG ATGTTTTACC TAAATCAAAG TACACATCGT TATCTATTCT    8880

TAAAGTAGTG CTACTTGTTT TGGATAGTAA GCACTCATAA ATACCGCCAC CGTTGCCGTC    8940

TGAGTCAACT ACATTCGCTT TACTCAATTG AATCGCGTTA GGTAATGCGG TTAGTCCGAA    9000

TCCCTCAATA ACGCCACCTG GATAAGTTCC ACTTACCAAC AAAATAGAAT AGTTTGTGTA    9060

CGGTTCAGTT AGATTGATTG TTGTACCTAC ACCATTTGCG CCACCGTCGA ACAATACCGT    9120

TGATTTATGT TCATTAGGAA CTGTCCACTG TTGCTCAAGT CTGCCGTTTG TGATTGATCG    9180

TGTGTAAATC TTTTTAGAGT TATAAGGTGT GAAGTTAAAT AGCTTGTTTG TATCATCTTT    9240

AACGAATACC GATAAATAAC CCTCATAACT TTCAACGCTA CCTGGTAAAT CCGGCACTCT    9300

TGTTGCATAG TAATTACCAG CAGTTAAATA TCCCAAATCG CCTTGCGCAT TATTTAAGTT    9360
```

```
GAATTTATCA TCTACATACT GCTTAGCTTG ATTTAAAGCG TTGTTAGACG TTTCTTCAAC      9480

AAATTGCTTA GTTAAGTTTC CATCATTCTT TTTATAAAAC GGGTACCATG TGCCGTAGAT      9540

TTTGTATTTT GTGTACTCAT CGTTTGAATC GTCTGGGTAC CATGTTGCAC GAGCAGTATT      9600

ATTATCAACA ACATAAACAA CTAACACACC AGATTTGCTT GATGTATAAG TTGATTCATC      9660

GAACGAAGAA CCGTCATCAA CACCATCTTG TCCAGGCTTC TCTAACGTGC CTATATCCGT      9720

CTTTTCTGGC GCATCTGTTG CATTAGTAAT ATGAATAATC CTAGATGTGT TAACTGCGCT      9780

TAAAACGCTA TCTATGGACT GCTCATACGA TTCAATTGCT TTACCGTAAT CATCTGTAAG      9840

TTTAGACTTT TGCCAATTCG TTGTTGAATT ACCTTTAACA AGGTCAGCGC CATTGATTTG      9900

TTGTTCAACT TCGTTAACAC GTTCAAAAAT CGCTTGCTCT TTTTCAACTA TTTTATCGAA      9960

TTCAGCTGTA ACAGCTTGTG TTGCACTAGT TTGCGTCGCA GTAATAGCTT GTATAGCTTC      10020

GTTTTGCTTG ATTTCGATTT GTTGAATGCC TTTTGTCGCA CTATCATTCA CTTTTGCTAT      10080

TAACGTTTGT GTATCAGCCA TATTTTGCTT TAATTGGTTA AAATCTTTAC CGACAGCTTC      10140

GATAGTATCT TGAATAGATT TGATATAAAC AAGCTTTGTT ATACCATCAA ACCCACTAAC      10200

TAAATCATTT TCAATATTGA AGCTAAATTG ACGTTCAACA ACAACATTAT TACTCCCGTT      10260

TTGTGTAAAG AATGCCTGAG CATGCACCTT GCCTGAATGT TTTAAAAATT CATTCGGTAT      10320

CACATACTGC AAACGCCCAT TAATTGCGTC TACTATCGTT AATTCGTCTG AAATATAAGC      10380

GCCTCTATCT ACGTTATAAT CATCGGTTTT TAAnACGATA GATGTTTTAA CATGTTCAGA      10440

ACTTATAGAT AAGGGTCTGT TATnCTTAGT                                      10470
```

(2) INFORMATION FOR SEQ ID NO: 21:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3647 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```
ATCAGATCTT GAGAATCGAG TTATTAAGTC TATCGAAGAC TTAACTAAAA TCCAACCATT        60

CATGCCTACA CAAGATTTTG ATTTTAAAAC TAAAGAAATT CAATCAAACA TTTCTGAAGA       120

AAGATTTATC GAAATGATTC AGTATTTCAA AGAGAAAATA ACAGAAGGGG ATATGTTCCA       180

AGTTGTGCCA TCAAGAATTT ACAAATATGC GcATCATGCT AGTCAGCATT TAAATCAACT       240

TTCGTTTCAA CTGTATCAAA ATTTAAAACG ACAAAACCCA AGTCCATATA TGTATTATCT       300
```

TCAAATTGTA ACAACTAATC CTATTGCAGG TACGATTCAA CGTGGTGAGA CGACACAAAT   420

AGATAATGAG AATATGAAAC AACTACTTAA TGATCCAAAA GAATGCAGCG AACATCGTAT   480

GCTAGTTGAT TTAGGACGTA ATGATATTCA TAGAGTAAGT AAAATCGGTA CCTCAAAAAT   540

TACTAAATTA ATGGTTATTG AAAAATATGA ACATGTTATG CATATCGTAA GTGAAGTCAC   600

AGGTAAAATA AATCAAAATT TATCGCCAAT GACAGTTATT GCGAATTTAT TACCAACAGG   660

TACCGTTTCA GGTGCACCAA AATTACGTGC AATTGAAAGA ATATATGAAC AATATCCACA   720

TAAACGGGGC GTTTATAGTG GTGGTGTTGG ATACATAAAT TGTAATCATA ACTTAGATTT   780

TGCATTAGCA ATTCGAACGA TGATGATAGA TGAGCAGTAT ATCAACGTAG AAGCTGGTTG   840

TGGCGTTGTA TATGATTCTA TTCCTGAAAA AGAACTGAAT GAAACGAAAT TGAAAGCTAA   900

AAGCTTATTG GAGGTGAGCC CATGATCTTA GTTGTAGATA ATTATGATTC CTTTACATAT   960

AACCTAGTGG ATATTGTTGC TCAACATACT GACGTCATTG TTCAATACCC TGATGATGAT   1020

AATGTGCTGA ATCAATCGGT GGACGCTGTT ATTATATCTC CTGGTCCAGG GCATCCATTA   1080

GACGATCAAC AGTTAATGAA AATCATATCA ACCTATCAAC ACAAACCCAT TTTAGGTATT   1140

TGTTTAGGGG CTCAGGCACT GACTTGTTAC TACGGTGGAG AAGTCATTAA AGGCGACAAG   1200

GTTATGCACG GCAAAGTTGA TACACTAAAG GTTATATCGC ATCATCAACA TCTGTTATAT   1260

CAAGATATAC CAGAACAGTT TTCAATTATG AGATATCATT CATTAATAAG TAACCCTGAC   1320

AATTTTCCAG AAGAATTGAA AATTACTGGA CGTACCAAAG ATTGTATACA GTCATTCGAG   1380

CATAAAGAAA GACCGCATTA TGGTATTCAG TACCATCCTG AATCATTTGC TACAGACTAT   1440

GGTGTCAAAA TAATTACAAA TTTCATTAAT CTAGTGAAGG AAGGATGAAA ACCATGACAT   1500

TACTAACAAG AATAAAAACT GAAACTATAT TACTTGAAAG CGACATTAAA GAGCTAATCG   1560

ATATACTTAT TTCTCCTAGT ATTGGAACTG ATATTAAATA TGAATTACTT AGTTCCTATT   1620

CGGAGCGAGA AATCCAACAA CAAGAATTAA CATATATTGT ACGTAGCTTA ATTAATACAA   1680

TGTATCCACA TCAACCATGT TATGAAGGGG CTATGTGTGT GTGCGGCACA GGTGGTGACA   1740

AGTCAAATAG TTTCAACATT TCAACGACTG TTGCTTTTGT TGTAGCAAGT GCTGGCGTAA   1800

AAGTTATAAA ACATGGtAAT AAAAGTATTA CCTCaAATTC aGGTAGTACG GATTTGttAA   1860

ATCAAATGAA CATACAAaCA ACAACTGTTG ATGATACACC TAACCAATTA AATGAnAAAG   1920

ACCTTGTATT CATTGGTGCA aCTGAATCAT ATCCAATCAT GAAGTATATG CAACCAGTTA   1980

GAAAAATGAT TGGAAAGCCT ACAATATTAA ACCTTGTGGG TCCATTAATT AATCCATATC   2040

ACTTAACGTA TCAAATGGTA GGCGTCTTTG ATCCTACAAA GTTAAAGTTA GTTGCTAAAA   2100

```
AAGCAACACT ATCTGGTGAT AATTTGATAT ATGAATTGAC TGAAGATGGA GAAATCAAAA    2220

ATTACACATT AAATGCGACT GATTATGGTT TGAAACATGC GCCGAATAGT GATTTTAAAG    2280

GCGGTTCACC TGAAGAAAAT TTAGCAATCT CCCTTAATAT CTTGAATGGT AAAGATCAGT    2340

CAAGTCGACG TGATGTTGTC TTACTAAATG CGGGTTTAAG CCTTTATGTT GCAGAGAAAr    2400

TGGATACCAT CGCAGAAGGC ATAGAACTTG CAACTACATT GATTGATAAT GGTGAAGCAT    2460

TGGAAAAATA CCATCAAATG AGAGGTGAAT AATATGACGA TTTTATCAGA AATTGTTAAA    2520

TATAAACAGT CACTTTTACA AAATGGCTAT TATCAAGACA AACTTAATAC CTTGAAAAGT    2580

GTGAAGATTC AGAATAAAAA ATCTTTTATA AACGCAATTG AGAAAGAACC AAAGCTAGCA    2640

ATTATTGCAG AAAATTAAATC GAAGAGTCCT ACAGTTAATG ACTTACCTGA ACGAGATTTA    2700

TCGCAACAAA TCTCAGATTA TGACCAATAT GGTGCAAATG CCGTGTCCAT TTTAACTGAT    2760

GAAAAGTACT TTGGTGGTAG TTTTGAAAGA TTACAAGCAT TGACGACAAA AACAACATTA    2820

CCCGTATTAT GCAAAGACTT TATTATAGAC CCGCTTCAAA TTGATGTTGC TAAACAAGCT    2880

GGTGCATCTA TGATTTTATT GATCGTTAAC ATCTTATCTG ATAAACAATT GAAAGATTTA    2940

TATAACTACG CTATATCGCA AAATCTAGAA GTGTTAGTTG AAGTACATGA TCGCCATGAA    3000

TTAGAACGTG CCTATAAGGT TAATGCTAAA TTGATTGGTG TAAATAACAG GGACTTAAAA    3060

CGATTTGTTA CAAATGTGGA ACATACAAAT ACTATTTTAG AAAATAAAAA AACAAATCAT    3120

TATTATATTT CTGAAAGTGG TATTCACGAT GCATCTGATG TAAGAAAAAT CTTGCATAGT    3180

GGTATCGATG GCTTACTAAT AGGTGAGGCG CTTATGCGTT GTGACAATCT ATCTGAATTT    3240

TTACCACAAC TGAAAATGCA AAAGGTGAAG TCATGATGAA ATTGAAATTT TGTGGCTTTA    3300

CATCAATAAA GGATGTTACA GCGGCCAGTC AATTACCTAT TGATGCGATA GGTTTCATCC    3360

ATTATGAAAA AAGTAAAAGG CATCAAACAA TTACCCAAAT AAAAAAGTTA GCGTCTGCTG    3420

TTCCAAATCA TATCGATAAA GTATGTGTCA TGGTAAATCC TGATTTAACA ACAATTGAAC    3480

ACGTATTAAG CAATACGTCA ATTAACACAA TACAGTTACA CgGCACAGAA TCTATTGATT    3540

TTATACAGGA AATTAAAAAG AAATATTCAA GCATTAAAAT CACTAAAGCT TTAGCTGCaG    3600

ATGgAAAACm TwATCCCAAA caTtAAtnAA tnTTAgGGGG TCCGTGG                  3647
```

(2) INFORMATION FOR SEQ ID NO: 22:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 5966 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```
CcAcCTTGAC CACCTTTACG TGGAATCTTT TCmCCTkGAG CAACaTCGaT AATaTATATT    60

GAAAgTCAAC AAGTTCTGGA CTAAATGTTG CTGCTAAGTT ATCGCCACCA GATTCTATGA   120

AAATTAGTTC TATATCGTCA TGACGTTCTA ATAATTCGTC TATTGCTGCA AAGTTCATAG   180

ATGCATCTTC ACGAATCGCA GTATGAGGAC ATCCACCAGT TTCAACACCA ATGATACGAC   240

TTTCAGGTAG AACTCCTGAA TTTACTAATA TCTTTTCGTC TTCTTTTGTA TATATATCAT   300

TTGTAATAAC GCCGATACTC ATTTCTTTTG AAAGACGTTT TACAACTTTT TCAATTAATT   360

GTGTTTTACC TGCACCTACA GGACCACCAA TACCAATTTT AATCGGATTT GCCACAATTA   420

TAACCTCCTA TGATATGAAA tTCTAACATT GaCGTTCTCA TGCGCCATTT GATTTAGTTC   480

TAAACCAGGC GCTGTCATGC CAAAATCTGC TTCTTTTAAT TCGAAAATCT GCTTTCTTGT   540

TCCTTCTATA TAAGGAATCA TGTGAGTAAC TATCTTTTGA CCAGCAGTTT GTCCAAGTGG   600

AATAGCACGA ACAGCATTTT GAGTTAAACT TGAAACATTT TGATATAAAT AGTAATCAAT   660

AATCGTTTCA ATATCTACAC CTAAATGATG GCCTAGCATA GTAAAACAAA TAGCTGGATT   720

TnACTTTGCT TTCTTATCTT GCATTTGTTG ATGATACCAA GCAATCCATG GGCTATtATA   780

AAGTTCTAAA GCCAATTTAA CCATGCGAGT CCCCATTTGT kTTGCACCAA CACGTGTTTC   840

TTTAGGTAAG TTTTGrACAr ACATCAGTTT ATCTATGTGT AATACTTTTT GTGTATCATC   900

ATTTTCCAAT GCATCATAAA CTAaACGCAT GGCTAAACCA TCAGAATAGG TAAGTTGCTC   960

TTGTAAAAAC ATTTTTAACC AAGCAATAAA AGTATGATCG TCATGAATTA TATTTCGTTG  1020

AATATATGTT TCAAGACCAA ATGAATGACT GAAAGCACCT GTTGGAAACT GTGAATCACA  1080

GAACTGAAAT AATCTTAAGT GTGTATGATC AATCATGAGA ATGCCCTATA TGTCTGAAAG  1140

CCTTATTAAC TTTACGGTCT TCTCGAACAT ATGGGATGCC TAAACTTTTT AATAAATCTT  1200

CAACTAAATA ATCATATTGT ACTAGCATTT CAGTCTCTGT AAATTGTGCT GGCAAATGAC  1260

GATTTCCTAA TTGATGGGCT ATATCTCCCA TTTCTTGCAA TGTTCTTGGT TGAATCACTA  1320

AAAGATCTTC TGAATTAACA TCCACAATAA TCATATTATG GTCATCTGCG TATAAAATAT  1380

CTCCATATTG TAAGTCAATA GGTTGTTTTA AACGAATGCC TATTTCAGTG CCATGGTCTG  1440

TAACGACTCT TTGAATACGT TTAACAAGAT CTGAATTTTC AAGGTATACT TTTTCGACGT  1500

GCTTTTGTTT TTCTGAATTT GACAAATTGG CAATATTGCC TTGGATTTCT TCAACAATCA  1560

TTCTATGTTC CTCCTAGAAT AAGAAGTATC TTTGAGTTAA TGGTAACTCA GTTGCTGCAT  1620

TACTTGTAAT TTTTTCTCCA TCTACATATA CTTCATATGT TTGTGGATCA ACGTCTAATT  1680
```

280

```
GACGCACCAT GCGTTTTAAA TTTAATGCAC GATTGATACC ATTTTCATAA GCAGTTTTAG    1800

ACACGAATGT CATTGACGTA CTTGTAAGGT TTCCGCCGTA TTGACCATAC ATTTTACGGT    1860

ACTTCATCGG TTCAGATGTA GGTATAGAAC CATTTGCATC GCCATTTACG GCAGAGTTAA    1920

TTAATCCGCC CTTTACAACT AATTCAGGTT TAACCCCAAA GAAAATTGGG TCCCATAAGA    1980

CAATGTCAGC TAGTTTGCCC GGCTCGATAG ATCCTACATA TTCAGAAATA CCATGTGTAA    2040

TTGCTGGGTT AATTGTATAT TTAGCGATAT AACGTTTGAT GCGATTATTA TCATTATGTT    2100

CAAAATCACC ATCTAAAGGA CCACGTTGTT CTTTCATGCG ATGTGCTACT TGCCATGTTC    2160

GTGTAATTAC TTCACCTACA CGGCCCATTG CTTGTGAATC GGAACTAATC ATACTGAATA    2220

CACCCATATC TTGCAGAACA TCTTCTGCTG CAATCGTTTC TTTACGAATA CGTGAATCTG    2280

CGAATGCGAT ATCTTCAGGA ATAGCCGCAT TTAAATGGTG AGTAATCATT ACCATATCTA    2340

AATGTTCATC TACAGTATTA TGTGTATAAG GCAAAGTTGG ATTTGTAGAT GAAGGTAAAA    2400

TATTTGAAAA TGCAGCGGAT TTAATTAAAT CAGGCGCATG ACCGCCACCA GCACCTTCAG    2460

TATGGTACAT ATGAAGTACA CGGTCTTTAA CAGCAGCCAT TGTGTCTTCC ATAAATCCTG    2520

CTTCATTTAA AGTATCTGCA TGTAATGCAA TTTGAACATC AAATTCATCA GCAACATCTA    2580

ATGCATGACT CAAAGCAGAT GGTGTTGCAC CCCAGTCTTC ATGTACTTTT AATCCAATTG    2640

CTCCGGCATT GATTTGTTCA ATGAGTGCAG TTGGATTTGT TGCTTGTCCT TTACCTGTAA    2700

AACCGACATT AATCGGTAAA CcTTCGGCAG CTTCTAACAT TCTATGAATA TGCCATGGAC    2760

CTGGAGTTAC AGTTGTTGCT TTAGAACCTT CTGAAGCACC AGTACCACCA CCAATATGAG    2820

TCGTAATACC ACTTTCTAAT GCGACCTCTG CTTGTTCAGG ATTAATAAAA TGAACATGAG    2880

TATCAATACC ACCAGCAGTG ACGATTTTAC CTTCAGCGGC AATGATATCT GTTGTTGAAC    2940

CTATAATAAT GTCGACATTA TCCATTATAT CTGGGTTGCC GGCATTACCT ATGGCGAAAA    3000

TATAACCATT TTTAATGCCT ATATCAGCTT TAACCACTTT ATCGTAATCG ATAATAACGG    3060

CATTAGAAAT GACAAGGTCT GCAACGTTCA CGTCATCACG TGTTACACGA GGATTTTGCG    3120

CCATACCGTC TCTAATAGAT TTACCACCAC CAAAAGTAGC TTCTTCACCA TAAACCGCAT    3180

AGTCTTTTTC TATTTGAGCA AATAGATTCG TATCACCTAA ACGAATGGAA TCTCCAACAG    3240

TTGGACCGTA TAAGCTCGTA TATTGATTTT GCGTCATTTT AAAGCTCATG ATCTTTTTCC    3300

TCCTTTTTTA TTCACGTTTT CAGCACCGTT ATCTCCGAAT ACACCTGCAT ATTCATCATT    3360

TTCATCAGTT GGGCGATAGA CACGTGACTC ATCGATAGGA CCATTGACCA TACCACGAAA    3420

ACCAAAAATT TTACGTTTGC CAGCATATTC AACTAATTGA ACTTCTTTTT TATCCCCAGG    3480
```

```
TTCGAAATCT AATGCTGCAT TTGCTTCATA AAAATGAAAA TGTGAGCCCA CTTGAATTGG    3600

TCGATCTCCT GTATTTTCAA CTTCGATAAC TGTTTCAGGA TGATGGTTAT TAATTTCAAC    3660

CTCTGTACTT TTTGTAATAA TTTCTCCTGG TATCATTTGA CTGCCTCCTT TAAACAATAG    3720

GGTGATGTAC TGTGATTAAC TTAGTACCAT CGGGGAACGT AGCCTCGATT TCGATATCTG    3780

TAATCATGTG TTCGACACCA TCCATGACAT CTTCTTTGTT TAGAATTTGT CTACCATAAC    3840

TCATTAACTC TGCAACGGTC TTACCATCGC GTGCACCTTC TAATAATTCA TCGCTGATTA    3900

AAGCTAATGC CTCAGGATGA TTTAGTTTCA AACCACGTGC TTTACGACGA CGTGCAACTT    3960

CCGCCGCCAC TACAATCATT AATTTGTCTT GCTCTCGTTG TGTAAAATGC AAATTAAAAC    4020

CCCCAATTTC ATATTAGATA CaATTTACAA AATTTATATT AATCCTAATT GTTGTGATAA    4080

ACAAGTAATA TACAAAGTTC AATGTGTAAT TAGAAAATTA TATTTTTAGC ATATCCGATA    4140

TTGAAGCAAA CAATCTAATC GAAAACAAAT AGTGGAATAT ATTTATGTAA AAACCAAAAT    4200

AGTTTTTAAT ATAACTTTTC ATAGAATAGT AGTATATTAA TGAGTAATGA TTCAAAGGAA    4260

AGGTGAAAGA TTTGAAGATA ATAGATGTGC TTTTGAAAAA TATATCTCAG GTTGTGTTAA    4320

TTAGTAATAA ATGGACAGGA TTATTTATCT TAATAGGATT ATTTGTAGCC. GATTGGACAA    4380

TTGGATTAGC GGCTATTGTA GGTAGCATCA TCGCCTATAC TTTTGCGCGT TTTATAAATT    4440

ATAGTGAGGC AGAGATTAAT GATGGGTTAG CTGGATTTAA TCCAGTGCTA ACTGCCATTG    4500

CGTTAACAAT CTTTTTAGAT AAGTCAGGAT TAGATATTGT TATAACAATG ATAGCAACTT    4560

TATTAACGTT ACCAGTTGCT GCTGCAGTGA GAGAAGTTTT AAGACCATAT AAAGTTCCGA    4620

TGCTGACGAT GCCTTTTGTC ATTGTGACTT GGTTTACAAT TTTACTTTCA GGACAGGTTA    4680

AATTTGTAGA TACATCGTTA AAGTTAATGC CTCAAAACAT TGAAACGGTT AATTTTAGCA    4740

ACAATGATAG AATaCATTTC ATTCAGTCAT TATTTGAAGG ATTCAGTCAA GTATTTATCG    4800

AAGCGAGTGT AATTGGTGGC GTATGTATTT TAATCGGCAT ATTGATAGCA TCAAGAAAAG    4860

CAACACTCTT AGCTGTTATA GCTAGTTTGT TAAGCTTTAT CATTGTAGCT CTATTAGGTG    4920

GTAATTATGA TGATATTAAT CAGGGATTAT TCGGTTATAA CTTTGTATTA ATGGCAATCG    4980

CACTAGGATA TACATTTAAA ACAGCGATTA ACCCTTATAT TTCGACTTTT TTAGGTGTGT    5040

TATTAACAGT AGTGGTGCAA CTAGGTACAA CAACATTGCT TGAACCGTTT GGCTTACCTG    5100

CATTAACATT GCCATTTATT ATCGTGACAT GGATTTTATT ATTTGCTGGT ATTAAACATG    5160

ACAAAGTAGA TGCTTGATAG TTAAATCAAA CCTAATATTG TTTGAATATC ACCTTAAACT    5220

ATACAGCGAA TTGTATAGTT TAAGGTGTAT TTTTATGGAT AAAATTAAGT GCATACTTAA    5280
```

GTGTTAAACT AGGAATAAAT AATTTATATT GTGTGTTGTG TGGGGTGACT AATATGAATG   5400

ATATGGATAA TTCCTTTTTA ATAACAACGG AAAATTCAAAG AAAATGGATT GAAAAATTCA   5460

AAGTAATTAG AGATACATTT AAGGCTAAAG CTGAATATAA TGATCAACAT AGCCAATTTC   5520

CATATAAAAA TATTGAATGG TTAATTAAAG AAGGTTATGG AAAATTAACG TTACCAAAAG   5580

CATATGGTGG TGAAGGTGCG ACCATAGAAG ACATGGTTAT TTTGCAATCA TTTTTAGGCG   5640

AACTTGATGG TGCCACAGCA TTATCTATTG GTTGGCATGT GAGTGTCGTA GGACAAATTT   5700

ATGAACAGAA ATTATGGTCT CAAGATATGT TGGAGCAATT TGCTGTTGAA ATTAATAATG   5760

GTGCATTAGT TAATAGAGCA GTTAGTGAAG CTGAAATGGG TAGTCCAACA AGAGGGGGAA   5820

GACCAAGTAC ACATGCTGTT AAAGCTGATG ATGGGTATAT TTTAAATGGT GTGAAGACAT   5880

ATACATCAAT GAGTAAAGCA CTAACACATA TTATTGTTGC TGCTTATATA GAAGAATTAG   5940

AAAGTGTTGG TTTTTTCTTA GTAGAC   5966

(2) INFORMATION FOR SEQ ID NO: 23:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 17310 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTGTGTCATC GCGAAATAGT TAGGGTCATT CATTAATCCT TTTGAACGTA TTTCATCAAA   60

ATATAACAAT TTCATTAGTA AAGGGGACTT GTTCAAACCA GCTATAATAC AAAATAGACC   120

TATAGTCACA CTGCTTATAA TATAAGAGGT AACGATCACT TTTTTGCTAT TACCTAACTT   180

AAAGATGATC ATCCCTAAAT AGAAATAAAT GACTACAAAT GCATATTTAA CTGTAGATGC   240

AAGAACTTCC TTAACCGTAA TAAATATCAA ATCATCAAAA AATaGCaAAC AArGCGTAAT   300

AATCATACGA TATGTATACA AAATAATGAm AAACTGTmAA AAATGATTTG CCTTTAATAA   360

ATGGTTAGCG AAAAACAGTA AATAAACTAA TATTAGTAAT GTGATAAAGT CAGCTATAGA   420

AACATTCACA CCGGCAATAA CCGAAGATTG CTGAATAAAA ACCGCTAAAC CGATAAGTAA   480

CAATGTTAGT AATTTACTAT TGTGTTGATT TTCCATTATA AACGTCTTCC ACTTCTTTAA   540

TCATTTTCTC CTCAGTAAAA CATTCTAAAT AACGTTTTCT AGATTGATTA CTCATTTTGA   600

TGTAATCACT GTCTATTAAA TATTTTTCCA GGACTTTAGC AATAGTTTCG GGTTGGTTGT   660

TCATCATACA TATACCATTA TCAGCTACTA ATTCTGAAAT ACCGCCAACA TGACTGGCTA   720

TTATTAAAAT AAACGTATCG TATTGTGATA ATAAATGACT CGCATTAATG ACATTGCCCA   840

AAAATGTGAC ATCATTTTCT AACCCAGCTT GTACAACTTG TTGCTGACAA TCATTTAATG   900

TAGGTCCATC GCCTATAAAT GTAAAATGCG CATGATTACT GTTATGTAAT TTCAATATCT   960

CTATTGCCGC GATTAGATTT TGTGGCAATT TTGGATAAGC AAATCTTGCA ATCATAACAA   1020

ATTGATGCTT TGTCGGGGCA TTAATCTGTA AATCTTGTTT ATTAGGCAAC ATTCCAACTA   1080

CTTCGCCAAT ATTGTTATGT GATTGGCTTT TTAGCGTTTG CTTAACAGCG GGAACATCTG   1140

CAATACCATT ATGTATTGTG GTTAATTTCA ATCGATTAAA TCGATATTTT AACGCTAACT   1200

GTTTATCGAA ATCTGAAACA CAAATAATGC TATCTGTAAT AAGTGACATT AATTTTTCGA   1260

TAACTAAATA TAGAAATTTT TTAGCTGGTT TAACACCCTC TGTAAAAGCC CATCCATGTG   1320

CAGTAAAAAC TATACGTGTG TCTTTCGATT TCGAAATGAa CTtCGCAATT CGTCcGACCG   1380

TtCCAGCTTT GGAAGAATGT AAATGGATAA CATCAGGTTT AATTTTCGAG AATAACTGTG   1440

CTAACACTTT GACAGCTAAA ATATCTTGTT TAAAGTCAAT TGGACCTACT AAATGTTCGA   1500

TAATAATTAC ATTAACTCTT GCATCTAGTT GTTCAATCAT TGGTCCATGA TTGCCTACAA   1560

TGACATAAAC ATCATTGTGT ACGCAAAAAT GGTTGGCGAG TTGAATGAGA TGTGTTTGTG   1620

CACCACCATT GTCTGCTTTA GTAATACAAT ATATAATTTT CAACTGTTAC AAACCCCTTT   1680

AATGCTATAC TTTCAATTTC TTAACATGGC TATCTCATCA GATGAATAGT ATTTATAGCC   1740

ATGCAAATCA ATGATGGCAC ATATTTCTTA ATGCCATTTG ATACTGTCTC AAGGGATTCC   1800

TCGTTATACT GTAACAATTG GTCACAATCT TTAAAATATA ACTTTTATTT GAACTTATTA   1860

AGTAAATTAA GACTACCTTG AGCCTTCCCC TGTAATAACA ACCATCAATG TTCTAATTGA   1920

TATATATAGT TCCATCATTA AACTACCTTT ATGTATATAT TTCATGTCAT ATTTCAGTTT   1980

TTGTTGCGGT GTTAAGTCAT ATCCACCTTG AATTTGCGCA AGTCCTGTTA ACCCTGGTGT   2040

AACAAGACAT CTTTGCTCGA AACCTATCAC TTCTGAACTA AATAATTCTA CAAATTCCGG   2100

ACGTTCCGGG CGTGGTCCAA TAAAACTCAT TTCCCCTTTA ACAACATTAA TTAGTTGTGG   2160

TAATTCATCA ATGCGTGTTT TACGAATAAA CTTCCCGACA TTTGTTATAC GATCATCATC   2220

TTTATCAGCC CATTGCGCAC CGTTTTTCTC TGCGTTTTTG CACATCGAAC GTAATTTGTA   2280

TATTTTAATT AATTTACCCA TCTTCCCAAC TCTAACCTGA CTATAAATAG GGTTTCCTGG   2340

CGAATCTATG ACGATAGCAA TGGCGAATAT AACCATAATC GGTAAAGTTA AAAATAATAA   2400

AACAATGCTT AAAATTAAGT CAATCGCACG TTTAATTGGG TAATAGCTTT TTCTCACTTC   2460

TTCTAGTTTG TCTAATTTTC TTTGATAGGC ATAACCCTTA TTATTATGGA CAGCTTCAAT   2520

```
AATTAAAGTA ATCCTTTAAA CCTGTTTCTA CTGTATATTT AGGAACAAAT CCTAATGCCT    2640

TTAAGTTAGA AATATCTGCA TAAGAATGCT TAATATCTCC TTTTCGTGCT TCTTTAAATT    2700

CATGCTCGAC TGATTTTCCA TATAATTCAC CAATAATACG ATAAACCTCT AATAAATTAG    2760

TAAAAGTGCC TGTACCAATG TTATAACCGT GTCCAATTGC ATCTTTGTGT TCCATAATTA    2820

AGCGTACAGA TTGAACAACA TCATATACAT ATACAAAATC TCTAGTTTGC AGTCCGTCAC    2880

CAAAAAATGT AAATGGCTTG TTATGCTCAA ATGAATCGAA CATCTTTGAA ATCACACCTG    2940

AATATTGTGA CTTAGGATCC TGTCTTGGCC CAAATACATT AAAAAATTTA ACAACCGCTG    3000

TTGGTATGTT ATATAACGAA CAATAATTTA ATGTCGTCCG TTCGCCGTAA TATTTATCTA    3060

TTGCATATGG TGATAATGGT AAGATTAATG ATTGATCACT TTTAGGCAAA TCAGGAAGAT    3120

CACCATAAAC AGCTGCTGAC GAAGCAAAGA TAAAACGTTT TATATGATTA TTATATTTTT    3180

TAATGATTTC TAACAATCTT AATGTTGCTA CGACGTTTAT TTCTTGAGAT AAGATAGGTT    3240

TCTCAACCGA CTCAGCAACA CTAACTAATG CTGCTAAATG AATAACATAA TCAAATTGAT    3300

ATGTCTTCAT GATTTGTTCA ACTGCATCAT ATTCACGAAT ATCTAATTCA AACACATGAT    3360

CGTCAGCCAA ACTTTTAATA TTTTCTCGTT TACCTGTTCT ATAGTTATCT AGAACATAAA    3420

CATCATAATC TTGTTGTAAA TCATCTACTA AATGCGACCC AATAAAACCA GCCCCACCAG    3480

TTATCAAAAC TCTTTCCAAA TCTTCCACCT CATTTATACA TTAAAAATAT ATCATAAAAA    3540

CATAAAGTAT TGTAAGCTTT TTATCGATAT TTTTTATTTA TAAAAATAAA ATGAGATAAC    3600

TTTGTGAATT TTTATTGAGA TAAATTAGAT AGTGGTGTTT TTGTGATGTT TTATAATATC    3660

TTGGGTGTGT TAATACTAAT AATGCTTTCA ACTGATGCAT TAGACTGTGA CATCATAACT    3720

CACTTAAGAA CTTCGCTTAT TAATTTTCTA CCAATACACT CCCTTCTAAG TGCACTAAAA    3780

AATCCTTACT GCTAAGTGAT TAAACTTAAC AATAAGGATT TATTTATCAT TAGTGGATGA    3840

TTATTAACGG AATCTCATAC CACCATCTAC AATAATTGTT TGTCCAGTAA TGTAATCAGA    3900

GTCTTTACCA GCTAAGAAGC TCACTACATT TGAAACATCT TCTGGTTGAG AAACTCTGCC    3960

CAAAGCAATC TGACTTGTAA ATTGTTCCCA ACCCCATGCT TCAGGTTTAC CTGCTTCTTC    4020

GGCTGTTGCC ACTGCGATAC TTTCCATCAT TGGTGTTTGA ACGATACCAG GTGCGAATGC    4080

ATTCACAGTA ATACCTTCAG ACGCTAAATC TTGTGCGGCT ACTTGTGTTA AACCTCGCAC    4140

TGCGAATTTT GTACTGCAAT ATAAAGACAA GCCTGGGTTA CCCTCAACGC CTGCTTGAGA    4200

TGTTGCATTG ATAATTTTAC CGCCATGATT GAATTTTTTA AATTGTTCAT GTGCGGCTTG    4260

AATACCCCAT AGCACACCTG CAACGTTCAC GCCATATACT GTTTTAAACT GTTCTTCAGT    4320
```

```
GCCAAATTGC GCGGCAGTTT GTCTTAcTGC GTTAAATACA TCATCACGGT TTGATACATC      4440

TGCTTTGATA GCAATAGCTT TTGTACCATC ACTTGATAAT TTAAGTGCAG CTGCTTTTGC      4500

CCCTTCTTCA TTGAAATCAA CAACTGCTAC TTTGAAACCA TCTTCCACTA AACGTTCTGC      4560

AATTTTAAAA CCAATCCCTT GTGcTCCGCC AGTTACTAAT GCTACTTTGT TGTTTGTCAT      4620

AAAGATCACT CCTCAAATTT CTTTCCTTTA ATTACATTTT ACTCCTCTTC ATTTGAATAG      4680

TACAACAAAG GTAGCTCCAT TTAACAAAT ATTCAGATAT TTAAGGTATA GTTAAACGCA       4740

CTACCATTAG TGATTGGCAA TGCGTTTAAA TGTCGTTTTA AAAGTTCTTA TGTTGAATAT      4800

TATTTTTTTA AGTCTCTCGA TTAGTTTGTC ATCAATCTTT TTTCGAGACA TGGTCTTTTG      4860

ATTCAATAGG CGGTTCCGTG TTATCACTGA CAACTTTAGT TGTAGCTTCA TCTTTATGTA      4920

TTTCTTCGTT AAATCCTTCA AGGTTTTTAG TCGTGGGATT TTTAACCTCA GGATGTTCCA      4980

TCATGTCTTG ACTATCAAGT TCCTTTTTAC ACGTGTCTTT ATGTGATGCT TGATTTGCGT      5040

TCCCTTTACT TTTTTGAATA GTGGTAGTAT CTGCTGCAGC TACTAATTTT TTTCTACCTA      5100

AAATAGATAT GGCTGAAACA AACCAGAGTA TTGCAGATAC AAAGTTGCAT AATACTAAAG      5160

CGATAATAGC CAATACAATT AATATGACAC CTTTTGAAAT CCTTTCTTTA AATAAGTCAG      5220

ATGCCAATAC GATGACAGGT ACGATTGAAA GTATAATTAC AAATATAGAA ATTATTGCCG      5280

ATATAACTAT TGTTACTATT AAATAATCAG CTCTGCTACC TGATAATAAA TAGAAAAGGC      5340

CGAAAATTAG TCCATAGCAA ATTACAAACC CACATAAAGT TATAGCCATG AGTACTATAT      5400

AAGCTATTTG AAAATATAAA CCTATCTTTA TGAATGATTT TTCTACATTT TTTTCCATGT      5460

CTATTCCCCA TTTATTTAAA ATTTATACTT TACCTTAAAT ATTCTCTTTA TTCTTTAGTG      5520

ATTTTATCTT TAGATTCAAA TTGATTCTCT GTACTTTCAA TATCAACTTT TTCATTTTCG      5580

TCTGTCGATT CATCTTTTGA GTATTTATTC CAAATCAGCA AAATACCACC AATCAGCCAT      5640

AAAATTGACG AAAGGAAATT ATATAAACAC AGTGCAATAA TAGCATAAAC AATAAAAAGT      5700

GCACCTCCGA TTACAGAGTA ACTTTCCATA TAAATCGCAG TAAAGATGGT TGGTAAAACA      5760

GTGAAAAGAG CCAATATTAA TCCTAATAAA AAAATTGTTT CGTAATCAGA TCCTCCAGCA      5820

ATATTAATAG ATATCATCCT AACAAAAACG ACACTAAAAT ATATTGAGC TACGATGCCT       5880

ATCCAAATTG CTATTTTTCC TATAATTGAG CTCATACTCA TTCCCCATTT ATTTAAAATT      5940

TATACTTTAC CTTAATATAC CTTATTTTAT TTAATTTTTA TATGCAAAAT ACAAAAATGG      6000

AGAACTTCAA TATTTATAAA ATATCAAAAG TTCTCCACAC TATATTGTTT TATTATATTT      6060

TCGCTATCAA TACGCTAAAT CATCATATTT CCCTCAACAT CACAGTAAAA CTATTGCTCC      6120
```

```
TTCCAATTGC GCAGTTGTTC AACATCATCA TCTTGTTTAA GTAATGCCAG TGGTACTTGA      6240

AGATTAAGAC ATCGTCCTGA AATATTAAAG CGTGTCACAC CTGCTGGCAC AGTTTCCCCT      6300

TTATGAACAA CCGCTTCAAT TTCCTTATAA CTCAATGGCT GATACTTCAT GAGTACATCT      6360

TGTTGAGAAA GACAAGGATA TGTACCTTGT GCAATTCTCT CTACAGAACA ACAACCACTA      6420

TAACTTGCGA CAACCTTTTC CCATACTTGA AAATGTGCTT CGCCTAAATC TTTTGTATAC      6480

AAATATTGTT CTGTATCACC ATGACACATT GTAATAAATG GCGCTTCTTG TCTTGTCTCA      6540

GTAGTCCATG GCAAGCGATG TTCTTGTTGT AACGTTTCCC ACCACACACC AAATGGAACT      6600

TTATGTTGCC ATGTACTAAT TGAATATTGT GTTTCATGGA TTTCTTGCAC TGGAACTTTC      6660

TTACATCCTA ACGCTTTCAA ACTTGTATAC CGATGCACAC CATCTATAAC CATATATCTA      6720

CCATGTTGCA TCGCTGTCAC TAAAATAGGA TGACGTATAA AATCATCTGC TTCAATACTA      6780

CTTTTCGTTT TTTCCAATCT TAAAGGTTCG AATGTTTCGT GAAGATCAAT CTTATCTACT      6840

GGTACCAATT TTAAATGTTC ATGAATATGA TTCAATAGTT ATTCATCCTC CTTTGTTTGT      6900

GTTAAATAAA TAAATTCAGG ATGTGGATGG CTTAAGAAAT CGTGATGTGA AATAGACCAT      6960

CCGTATGCAC CTGCATATTT GAAAACAATA ACGTCGCCTG TACTGATTGC GTCTATCTGT      7020

ACTTCTCTAG CAAAGACATC TTTCGGTGTA CATAATTGAC CGACTAACGT TGTGTCCTGT      7080

CTCGAAATTG AAACTTTTTC AAATGAATAT GGATTGTCCT TATAGCGATA AATGTCAAAA      7140

GGATGGTTAT GTTGCCAAGA TACCGGCAGT CTAAATTGTT GCGTACCTCC TCTTAATATG      7200

GCATACCAAG CACCATGTAC TTTCTTAATG TCTAGCACTT CTGTCACATA GTAACCAATA      7260

TGTGCCACAA TAAAGCGCCC ACATTCAAAG TTCAATGTCA CATCTTCCAT TTCTTGCTCA      7320

ACGATAAGTG TTTTAAAACG TTCTACAAAA TTATCCCATT CAAATTGGTT AGTTAAATCT      7380

GCATAGTTAA CGCCTATGCC ACCACCAAGA TTGATATGTT TGAGTGGAAA TCGATGTTTT      7440

TCAGACCATG CCTTTGCTTT TTTAAAATAA AGTTTCACTA CATCGACATG TAAATTCGAG      7500

TCTAAATTGT TAGAAATAGA ATGAAAATGA AATCCATCTA GATGAATCTT TGGCATTGCG      7560

AGCGCAgcTT cAATGACATC ATCAACTTCG TCTTCAGAAA TACCAAATTG TGTTGGGCGT      7620

CCTGCCATAT GCAACGTTGC ATTGGGAAAT GGTCCTGCTA AATTAACACG CAATAAAATG      7680

TGTTGTGTCT TATCTTCATC TTCTAAGATG GCATTTAGCC GTTGTAATTC ATGCATACTT      7740

TCAACATGAA TACGCTGAAC ACCTTCACTT ACTGCATATC TTAGTTCCTC GTCTGTCTTA      7800

CCAGGGCCAC CAAAAATAAT ATGATTTGCT GGTTTAAAAG CAAGACCTTT TGCTATTTCA      7860

CCTTGAGATG CAACTTCGAA TCCTTCAACA TACTGACTAA TTGTATCTAG GATTTTTCGT      7920
```

```
TGTTGCAAAT GATGTTCCAG TCCGACTAAA TCATAGATAT AATGACAAAC TGGATGAGAT        8040

TGTGCTTTTA ATTGTTCAAT AACAGGTTGA ACTATACGCA TTAGCCTTCA TCCCCTTTCT        8100

GTTTAGACGT CGCTAGAGAT GCACTTAAAT GGCGATATAT TTTTCCGCGA TCATCACCTA        8160

AAATAAATGT TTGTACACCT TGTGCCTGCC ATTTTGCAAT ATCTTCATCT TCACGTGGTA        8220

ATGCACAAAA ATGTTTACCA TGTGCATTCA CAACTTCAAA AATATGTTGA ACATGTGATG        8280

TTACTTGATC ATCACGCGTT TGCCATGGTA TGCCAAGTGA CTGCGATAAA TCTGCGGCAC        8340

CTTCGACTAT CATGTCTAAA CCTTCGACTT GTGCTATATC GTCAATGGCC ATAACCCCTT        8400

CAACATCTTC TATCATGGCA ATCACCATAA TATGCTCATT AGCCATCTCC ATTGCATCAA        8460

GTAATGGTGT ACGTCCAAAT CTTGCCATGC GACCACCATT CAAACTTCTT AATCCTTGCG        8520

GGTAATAACG ACTTAATTTC ACAATATGCT CAACTGTCTC ACGATCTTTA ACGTGTGGCA        8580

CAATAATACC TCTCGCACCC ATATCCAACA CTTTAATGAT ATCTCTATCT ATCACTGCAG        8640

TGACACGTAC AATTGGTATA ATATGCGCTG CTTCAGCTGC ACGAATTAAA TGCGCTAGTG        8700

TCTCATCATT AATCGCCACG TGTTCTGTAT CAATCACAAC AAAGTCATAC CCGCTTGCTG        8760

CGATAACCTC GATCATCAAT GGGTCCGGTA TAGAATTAAA AATGCCATAA ACTGAATCAC        8820

CATTGTTTAA TCTATGTTTC AGAGATAGTT GTTGCATCAT TGATACCTCC TACACCTAAT        8880

GGATTTGTAA CATGATGAAT TCTTAACTCG GAGTCACTTA ATAATCGACG TGTCGTTAAC        8940

TTTTCAACTT GAATCGTAGG TTCAAACAAA TCGAAATGTT GATAGTTATT CAACTCTGGA        9000

AATGCTTCTT GATACGCCTC GATGATGCCT TTAACCCATT GCCATTGCAG CTCCTCATCG        9060

ATACCATATT GCTTTTCAAT AAATAAGATG ATTTCGGCGA TATTAATAAA GAAAAATGCA        9120

TCATGTAAAA AGTCGCGTAC TAAACGTTCG TCATCTGTTT CAATAAATGA ATTACTATTC        9180

ACTTTTTTAT GTGCTTCTGG CATTGGCTTT AATGTCAGGT GTGAAGCAGC TTCACTTAAA        9240

TGCtCACGCT TAAAACGAAC ACCATCATGG AAATCTTTTA AGGCAATACG TGTAGGCCAA        9300

CCATTTTCAT GAATGAGCAT CATATTTTGT GCATGCGATT CAAAGGCAAT ACCGTGATAA        9360

TAAAGCATAT GAATCATTGG ACGAATCGCT ACAGCTAAAA ATTGCTTTGT CCAAGCTTCA        9420

GAACCATATT GTTTAATCCA ATTTTCAATG AATGGTACAC CATCCTTATC ACTTGCATAA        9480

AGTGCATTAA ATGGTATCGC ATCCTCTTCA TCGATTAACA TATGATATAT ATTTTCACGC        9540

CATATAACAC CTAACGCACC ATAAACTTGA GTTTGTTTAT AAGGCGAAAG TTGTGTATTT        9600

AAATAAGACT GTCCTAAGAC TTCCCCTAGA AAAACTGTCT TTAATTCATC TTTTAAATAC        9660

ATATCTTGTT GCTGTATCTG CTTTAACCAA TCCGTAATTT GCGCTGCATT TTCAATTGTA        9720
```

```
TATTTTGTCG TGTCTATTGG CGACATCGTA CGAATCGATT GTTGAGGGTG ATATAGCTCA      9840

TCACTTTCCC CTAACCATAG TACTGTGCCA TTAAGCCTTT CTTCAGCCAA ATCAACTTGG      9900

ATGACATGTT CAAACTGCCA TGGGTGTACA GGTATCATCT CAACATCATT TACATGTTTG      9960

CCAGATGCTT CAATTTGCTG TACAAAATGT TCATAAGTCT TATCGCCAAC TTGTTGACGT     10020

AACATTTCGT TAACTACAAC ATTTCTTGAT ACCGTCGTTT CTACTTTATC TTTGTCGATA     10080

GCTAACCACT GCAGTTTAAC GTTTGGTACA AAATCAGGAC CAAATTTCAA ATTATCACTC     10140

AACGTAAATC CTAAACGTGA TTTGTAACTT GGATGATACT GATGCCCTTC CATCGCATAA     10200

AATTCATAGT CGTTAAATGT CTCAGGTGTT GCTGGTGGGT TTGATTCTCG ATACTGCATA     10260

CTTTGCGTAT CTTTTAATTC TGTCTGTAAT AACTCGACAA TAAATTGTTC TAGCTTTTCA     10320

TCATTTTTAG GAAATGTAAA TACAACCTCT CTCAATAATT GTGTATAGTC TGTTGTTGTA     10380

TCTGCCTCAT CTCCTACGAC ACGCTCAATT GGTGATGTGA TACGTATACG ATCAAAGCTA     10440

TGTGTCTTTT CAGCAGTAAA ACGATACTCT GAATCATGTC CTTCTATTGT AAAATGACCG     10500

ACACCGTCTT GATATGACGC TTTATACACA ACAATATTCT CATAAATAAG TGATGATACC     10560

AGTTGGTGCA TCACTCTAGT CTTTACACGA TTAAGAATTG TTTGATTCAC AATACGATAC     10620

CTCCTTGTTA TGACAAATTG GATTTGGTAT ATGTGTATAA ATAGGGTTTG CACCACAATC     10680

ATTCAATTTA CTCATCAAAT TCGCTTTAGC CGcAATGGTC GGCGTTTGAT ATAAATCTTC     10740

TACACAGTCA ACAAATACTG CGTTATTCGC GTATTCTTTT TTCCAAGTCA TAAGACGATG     10800

CGCTACAAGT TGCCATAACA CAACTTCATT TCTAGTCGCT TTACCAATAG TTGATACTAA     10860

ATGTCCTAAG TGATTTACTA CAACGTAATA TTTAAGACGA TGCCATGCTT CATCATGTGC     10920

ATATACAACA GGGCTTGATG CTGCCACAAC ATTTGGCACA AGCTGTTTTT CAGTAGCAAT     10980

CGTTCTAGAT AGACAAATGC CTTCAAGATC TCTGACAAAG CATACGTCGG GTATGCCATC     11040

TTTTAATTCA ATTAATGTAT TTTGTACATG TGCTTCTAGA CTAATGCCTG TGTTACTAAA     11100

CAGCTTTAAT ATCGGCAATA ATGTACGATT CAAATAACAT TCAAGCCATG CTTCTGGTGC     11160

TAAACCACTT TGCTCAATCA CTTGTGATAA CTTAGACATC GGTGAATCAG GCATCGTTTC     11220

AAATAATGAC GCCAATACAT GAATATCTTT ATCAGCATGG TAATTCGGTA TCCCTTCACG     11280

AACAATCATG GCACTATTTG TTAATAAATC CATTTCAGGT TCAACTGTTT GCCCTAATGG     11340

ATTCGGTAAC AATGCACGAT ATCCTTCTTC AAACATCAAT TTAAAATGGG GTGTTTCAAC     11400

CTCATCTTTG ACTGATGCGA TAACTTGCGC GGCATCAATT GTCCGTTCAA TCTGTTCAAG     11460

GTCATTCGTA CGTATAAAAT TAGTGATTTT AACGTGTATC GGTAATTTTA AATAAATGTT     11520
```

```
GCCAAGGTCT TTTATTAAAC CTTGTTCACT ATATTGCATA TACTGTGGAT GCTGTCGCAA      11640

CACATTGATT TGATAAGGAT GTGTTGGTAA TAAAATAAAA TCTTTGGGTA TCTCTGATAT      11700

ATCTATGTCT GCTAATTGAT ACAACACTTT CTCAACCTGA TCTTCTTTAC CTTCTACATA      11760

GCGCGTGAGC AGAACATCTT GATGCACAGC TAAATAATGC AATTGGAATG ATGTATGACA      11820

TTCGGGTGCA TATTTCTCTA AATCTGCTTC TGAAAACCCA CTTGCACTCT TAGGAGTCGG      11880

ATGAAATGGA TGACCTAAGT ATAAAGATTG TTCTGAAACG ATATAACGAT CCTCTACGTA      11940

GTCTATTGTG TTACTTTGCA AATAACGTGC CGTGCGATGA ATGCTATTAT CGATGTCAGA      12000

CATAATTTGC GCCATATGTT GTTGCACTGC CGTTTGATTA TCTGCACTTT GAGCCATATG      12060

TTGCAAAATA CGCGCAATTG CTTCTTTATA AGTTGTTATT TTTTTACTTT TTCCATCGAT      12120

AAGCCATACC TCTGGATGAT ACATATGATG CCCCATCGCA GACCAATAGC GAAATTCACC      12180

CGTTAAAGTT TCGAGCTCTG ATAATTGTAT AGACCATTGA TGATTTTGAG GTGGTACTTG      12240

ATATAAATTT TCTTCTCTAA AATATTCATT TAAAATGCGT TCGATAGCCG CATACGCTGC      12300

ATGTTGTATT AATTCTTTAT TTTGCACTTT TTTGTTTCAA CTCCCATAAT TTCATTAATG      12360

TGTGATCGTT GATTTGATTA GTGATGGTTG AACAAATTAA AAATAAACTA CTTACTGCAA      12420

ATACTACGCC CATAACGATA AACGTAGTAG CTGGTGTAGT ATAACTTGTA ATGGCAGCGC      12480

cACTaAGACT GCCAATAATT TGACCAACAA CTAACATACT GTTCGTCGTT CCAACAAATG      12540

TGCCTTTAAG TTGTTGATGA CACGCATTCA CGACAACAAA CATGACACTT TGAATCAATG      12600

CACTATATGT TAATCCTTGA AGTATTCTTG CAGCCATTAA AAACTCTATA TTCGTCGCTA      12660

AACCTTGCAG TATCGCACTA CAACCACATG CAATCGTGGC AAATATATAT ACTGATTTAA      12720

CATATGATTT ATCATTAAAG CGTCCCCATA AAGGCGCGCT TAATATCGAA GCCGTCCAAA      12780

ATGCGGACTG TAAAAATCCA ATCACACTAC GGTCATCTAT CGCTGTATGA TTCACTGATG      12840

AAGCAAGTGG TGATAATGCA GTTAGCATGC CATACATAGC AAAGTTTGCT AAAACGCCAA      12900

CGATAATAAA TCGACATGTT TGTTGTGTGC ATAATAGACA TTGAAATGAA CGGCGAATAC      12960

CTTTATTAAT ATTTGGTGTT TGTGATTTTG GCATATGTGT CGTTCAATC AATTTTAATG      13020

CACCGAAAAT ACAGACAATA AAAGTAATAA CGGCAATACT CATCAGTAAC GCACTAAAAC      13080

CTAATATCGA AGCTGTAACA CCGCCAATTA ATGGCCCCAC AAGAGACCCT GCGCTGACTG      13140

AACTTTGCAG TCTTCCTAAT ACCTTTCCAC GATCTTCAGC TGGCGCCTCT GCACTCGCAA      13200

ACGCACTTGA TGCATCAACA ACACCACCAA ATAGTCCCTG CAATAACCTC ACAAGTACAA      13260

ACTGTAATGG TGTCGTACAC AATGCCATTA AAAATAAGCA TACCGCCAAA CCAAGTAACG      13320
```

```
CtATCATCGT CGTTACAGCT GGAGCAGCAA TCGCTATACC ACTCCACAAC TGTATTTCTA   13440
CGACTGATAG ATTTTGTAGT GATGCCATAT AAATTGGCAA TAATGGCACA AGTACTGTCA   13500
GTCCAGCAAT CGCTATAAAC TGACTGAGCC ATAAAATGCG AAAGTTACTG CGCCATATAG   13560
ACTGATTAAT CATATGTCAC CATTGGATTT GGTACGGTAG TTAAACCTGA AGGCATACTA   13620
CCTCCACCAC TATCACGTTG ATATAGCAAT GGTAATAAAA TTTGTTTGAA TGGCCACGTC   13680
TGTTTATCAA ATAAAATGTG TCTGACAGCT AGCTGATCAG TTGTAACCCA GGAAATAGTT   13740
GCCACTTCAT TTTTTAAAAT TTGTTTTAAC AACGACATAA GTTCATGCTC ACTTACACCA   13800
AATAAATCTT GAATTGCATC AATAATGGCA TATAGATTTA CCGATACAGC TAATGTTTGA   13860
AAATAAGCAA AGAATGTTTC CAAATCCTCA TTAATTAGCG TATTAGGTGT ATCTTCTCTG   13920
ACGACATACT TCGGCAATGA AAGCTGATGT GCTGTTAGCC ATGGTTTATA AATTCTGACA   13980
GTATCATGAT CACGTAACAC GCATTTTTGT ACACGTCCAT CTTCAAATGA CAACAATATA   14040
TTTTGACCAT GCAACTCTGG TAATGCGCCG TATTGCATAA ATGATAGTGT TACCTTTAAA   14100
AAGACTTGCG CGATATCTTC AAATAACGTC ATGACATCAT TTTTAGAAAT ATTATCTTTT   14160
CCACAAATCA TTTGATATAA AGTGCGATCA TTTGCCGCGA GTGCTGCCAT TGACACTAGC   14220
TGTTGCGTAT CATTTTTGGC TAGCACTTCG GGATACTTTC TTAGCTGAAC AGTTAGATGA   14280
CCTAATTGAT CTTTGAAAAT ATCATTATCT TGACCCATAT ATGACCACCA AGCTGTTTCA   14340
TCACAAACCA TGACATACTT AGCTAGTGCT TCATCTTTTT CTATAAGCTG ACGTAATAAT   14400
TGTTCTGCTT GTTCTCCGTT TTTCATGTAA CGCGTAGGCG TTAGCCTTAA TGCGCCTAAT   14460
GACTGCATTG CAAATGGTAC TTTGACATGG TTATACGGTG CGCCAATATC AATTAATGAA   14520
CGCATACTTG AAGACGACAG ATAATCTCCA AATTTTAACG GTAATAGTAC AACCAACTTT   14580
TCACTAATCT CTTTCGCAAA GACGTTCGGC AGAATATGCT GATATTGCCA AGGATGTACC   14640
GGAAATAGTA CATAGTCATC TATTGATAAC CCTTGATCAT TTAACATGTC TGTCGCTTGT   14700
TCTTTTATAG GTACTGTCAA ATTTTCTAAT TCATCGATAT TTGCAGTATC GCCATGAATC   14760
ATATGTGTCT TTTTAACTGC TGCAACCATT AAAGGAAATG ATTGATTTAA TTCAGCTTGA   14820
TACACTTGAT AATCCGCTTC TCTTAATCCT CTTTTTTCTT TAGCTAATGG ATGAAATGGA   14880
CGATCTTTTA AACTTGCAAA CTGCTCTGAC ATCACAAAAG GATGTGACGC TAAATCTAAT   14940
TCTGATAATT GTTTAGCAAG CTGTGTGGCA GCAGTAGTCA GTCCTTCTTC AACGCGAGCC   15000
ACTTCCCATT CATGACTTAG ATCACAATTC ATATTAGCAA TTGTTTGCCA AAATTCAGCT   15060
GCCGTTAAAG GTTGCTTAGA CACCCTTCCC TCTATCGTAA TTGGTTGTGA ACTTTCGTAA   15120
```

```
TATATCAAAA GCGTTTGTCC GTTTTCTTTA GTAATCTCAC TATTCGATAC AATTCCGGCT      15240

ATATCTTCAA ATAATAATGC ATCAACTAAA TCTCTTAATA TTATCGCTTG TGCTGTATTG      15300

ACTGCTGTAT GATTCTGCAA TGTTCAGACA CCTCGCATTC TTAATATAGG TTCAATGTTG      15360

TCCCAATATT TTGTTGTTGT GCCTGTTGAT AAATAAAATA AGCACTTGAA ATATCTTCGA      15420

TAGCCATACC CATCGGATTA AGTAATATGA TCTCATCATC GTCTTCACGT CCTGGTATGT      15480

CACCTGTCAC AAGTTGTCCT AGTTCAGCAT GAAGAGCTTC TTTGCTGAAT TTACCTTCTA      15540

ACACCAATTG GTTAATAGTT TTCTTTTCTC GATTACATTG TGACCAGTCA TCTACTACGA      15600

CTTTGTCAGC TTTAATAAAG ACTTCTTTAT GCACATCCAT GATAGAAATG TTGCTAATAA      15660

ATGCACCCTT TTGTAACCAA TCATATTCAA TGTATGGTTG ATCCGTTACG GTACATGTAA      15720

TGACTACTTC ACCATTTGAT ACTGCTTCTT TAGCATTTTC TGTCGCAATA AAATTAATTT      15780

CCGGACGCTG TTGTTGCCAT CTATCAACAA AGCGTGCACA TGCTTCAGAG AATTGATCGT      15840

AAACAAACAC GCGTTCAATA TGATCGAATT GCTCTAACAT ACTTTGTAAT TGCTTGTCTC      15900

CGATTAGCCC GCATCCAATG ATTGTTAAGT CTTTAAATCC TTTTTTAGCC AAATGCTTTG      15960

CTGCAATCAC TGAAACTGCT GCAGTACGCA TACTACTAAT TAAACTTGCT TCCATAACTG      16020

CAATTGGATA ATTCGTTTCT GGATCATTCA AAATAATGAC GCCACTTGCA CGCTCCATAT      16080

TACGTTTCGA TGGATTGTCG TGCTTACTAC CTATCCACTT AATACCTGAA ATTGCGTGTT      16140

CACCACCGAT ATGACTTGGC ATTGCAATAA TTCGATCTGC GATGTGTCCA TTTTCAGGAT      16200

CCtGTCTTAA ATACGGCTTA AGCGGTTGTA CAAAATCATT GTGCGCATGG GCTGTTAATG      16260

CTTCTGTTAA TGCGTCCACA TAAACTTGTG AATGATTACC TCCCGCTTGT TCAATATCTG      16320

ATCTATTTAA ATACAACATC TCTCTattCa TTCTGaTTTA ACTCCTTGTC TTGATTTCAT      16380

TTTTTCTAAC CATGTATCTG AATAAACTAA ATCTAAGTAA CGATCGCCTC GATCTGGTAA      16440

AATCGTGACA ATTGTTGCAC CTTCTTCAAT TGACGTTATC AACTGCTCAA TCGCTGCAAT      16500

AATCGAACCT GTTGAAcCTC CGGCAAATAT GCCTTCATAA TCAATCAGTT TTCGACAGCC      16560

CAAAGCAGAT TGATAATCAT CTACATGGAT CACTTGATTA ATTTCTGATC TATTCAATAT      16620

TTCGGGTACA CGACTAGCAC CGATACCAGG TAATTCTCTA TTAATAGGTT TGTCACCAAA      16680

AATGACTGAC CCTTTCGCAT CAACAGCAAC AATTTGTGCG TTTGGATGCA CTTCTTTTAT      16740

TTTTCTACTC ATACCCATAA TGCTACCTGT CGTGCTGACT GGCGCGACAA AATAATCTAT      16800

AGGTTGCTTA ATTGTTTCAA CAATCTCTGT GCCTGCACCA TGATAATGGG ATTGCCAATT      16860

TAACTCATTC GCATATTGAT TAATCCAATA TGCATCGTCA ATAGTGGCTA ACAGTTCTTG      16920
```

```
TACATTGGCA CCATAACTTT TAATAATTTT CAAATTTGTT GGTGATATTT TAGGATCAAC    17040

AACACACGTG AGTTTTAATC CCTTGATTTT AGCTATCATT GCCAACGCAA TGCCTAAATT    17100

ACCAGAAGTA CTTTCAATTA AATGTGTATT CTCAGTGATT AAACCATGTT TAATACCATG    17160

TTCAATGATG TACTTGGCAG GTCGATCTTT CATGCTGCCT CCAGGATTCA TATACTCTAA    17220

CTTTGCAAAC ACTTCATGTT TCGGAAATAG TTGATGAAGT TGAACCATAG GTGTTTGCCC    17280

TACAGAATCT AACAATGAAT CGTGCACATG                                     17310
```

(2) INFORMATION FOR SEQ ID NO: 24:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 5423 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

```
ATACTAGTAA GCGCATCGGT TATTGACATC GAATTCAACT TTAACAGTTT TCATGTTCGG      60

TGATGTTTCa ATAGAATGTG TGTGTTGTAC TTGCGCATTT ATATTTCCAC CTAAATTACT     120

TAAGTTTCCT GTAATACTAG AAATGTCAGG TGCGTTTAAT GTAGGTTGAA ATGCATCAAC     180

TACTTTATCT GCAACATTAG AAACATTACG GATAACTTTA CTTGAATGAT TATCTATACC     240

TTTAACGAAA CCTAACATTG AATACATACC AACATCCATG AATTCACGTG AAGGTGAGTG     300

AATACCTAGC GCTCTTTTGG CTGCATTTAA AGCACCTTTT GCTACACTAG CTGCTTTTTC     360

AGCTAAGTCT CTAGCCATAT TACCAATACC TCTCATCAAA CCACGGATCA TATCAGCACC     420

TGCTGATACA AAGTCATCCA CAAAGCTTTT AACTTTATTT ACTGCATTTG TCATACCTTG     480

ACTAACTTTG TTTACAACAT TAACGAATCC TTGAACAACT CTATTAACAA rGTTAATTAG     540

CGTACtTGTt ATAGTAGATA CCCaTnGCAT ACCTTTAGTG ACmATGAAGT TCCAAGCTTG     600

AGACATTTTG TCTGATATAG TTGAAACAAC TTGTGTGAAT ATGCTTACAA CTTTATTCCA     660

AATTGTCGTT AATATACCAG ATAAGAAACT CCAAATCGTA TTCCATATAT TAGAAATAAA     720

ACTCCATGCC GCTTGTAACG CAGTAGATAT AGCTGTAGTG ATAGCGTTCC AAACCTTAGT     780

TGCCACAGTA ACTATAGTGT TCCACAACGT TTGTAAGAAC GTCCAAATAG CGTTCCAAAT     840

TGTTATTGCG ATAGTCATAA TTGTGGTAAA CACTGTAGTT ATTACAGTGA CTAACAAATT     900

CCAAATCGTA GTAGCGATTG TAATTATCGT ATTCCAGATT GTACTTAAGA ACGTCCAAAT     960

AGCTGTCCAT ATCGTCATAA CTATTGTCAT TATCGTCGTG AAAACAGTTG TAATGATTGT    1020
```

```
ATAAGCGACT ATTTGATTCC AAACAATCAT TATAAAATTG TAAACATTCG ATACTGCTGT   1140

AGTGATAGCT GTTAAAATAG CATTCCATAC AACCGAAGCT ACAGCTTTTA ATACATTCCA   1200

AACATTAACC ATAAACGTTT TTATCGCATT CCAAGCATTT ATAATAAAGT TTCTGAATCC   1260

TTCATTTTTA TTCCACAATA AAACGAATAT AGCTATTAAT GCAGCAATTA CACCAATTAC   1320

TATTGTTATT GGACCGCCTA AAATACCAAA CACAGTTACT AGTCCTGTGA TAGCATTTCT   1380

AATTAATCCA ATCTTACCGA ATAACAATTG GAATATAACT GATATAATTT TTAATGGTCC   1440

TTTTAATAAC ATGAACGCAC CTTTTAAAAT TGTTAATCCC GCTCTTAATA AACCGAACTT   1500

ACTTACTAAT GCAATGrTTC TACCTATTAA TCCGCCACCC ATAAAGTTAG ATACAGCAAG   1560

AATAATCGGT ATTAAAAATC TAAATGCACC AACTAAAGTT ATAATGACAC CAACTAATTG   1620

TGCTGTAGCT GGATGCGCCT CAAACAAGTT AGCTATCCAA CCAGTTATTG CAACTGCAAC   1680

GCGTAATACT GCACTAGCTA TAGGAGCCAT TGCTGTTGCG AATGCArmTA ATCCTCTTGC   1740

GATGTTTCCA ATCAATTGCA TTATTAGTGG TCCATTTGTT TGTATATAAC TGACAAAGTC   1800

TTTAAACCCT TGAGATTGTC CTACTTGTTC AGACCATTCC CTAAACTTAG CTGTCATTTG   1860

TTCAAGAGAT TGGAATATGC CAGTTGATGA TCCGCTGAAT GCATTCATCA AATTGTTAAT   1920

TCCAACGAAA ACATTTTTGA AAATATTACC AATGATAGGT AAGTTTGTTT TTGTGTATTC   1980

AATAAAACGA GTTATCGAAT TTTCTCCAGC TGCACTATTA GCCCAGTTAG AGAAAGATTG   2040

ACCTAATCTA TCCAACCAAT CAGCCGACCA TTGAAACAGT GGTGCTAATT GCGTGAATAC   2100

ATTGACTAAT CCGTCACCAA AACCACCTGC AGCACTTAAT AGCTTGTTAA ATACCGAAAC   2160

ACCCGTTGTA TTCATCATAT TAAAGAATCT TGAAGCTACA CTGCTATTTT CAGCCCATTT   2220

AAGCACGCTT TGAGACGCTT CTTCCATTCC TCTTGAAATA CCACTAAAAA ACGGTTGTAA   2280

GCTCTGCATT GCAGTTTTAA CAGTATTTAA ACCATTTGCA AGAGTTGTGA AGATAGCGGA   2340

TTGATTTTGC TTTATAATAT CAGTCCATGC TGACTTTACG CCATCTAACG CTTTTTTGTA   2400

TTCGTTGTT GCTGAGCTAG CTTGTAAAGT GCCATCATTA AGCATCTTTA TAGCGCTGAT   2460

AGCCATTGCG CCAAACGCTA CAAATCCTGC TCCCGCTATT GCTACGGCAC CACCTAAAGC   2520

AAGTACACCA CCAGTTAACA CTTTGATAGC GTTTAATAGC GCAAATACTA CAGGTACTAC   2580

GCTCGCTATT ACAGGTATTA AGATACTAAA AGATGATGTA AGTAATCCAC CAACCATATT   2640

AGAACCTACA GTACCGAACA CACGGAACAT ATTAGCTAAA TTCCCCATCT GTCTTTGAAA   2700

ATTGTCATTT GCTTTTATTA TGTAGGCATA AGCTTTCTTT AAACCATTAG TATCGACATC   2760

TACCTTTGTT GTTTTTTTGT TCGGCAATGC GTCTAATGAT TTTTTAAACG CATAAATAGT   2820
```

```
AAGTTCTTCT TTAGTACGTT TGATTTTAGA GTTAGCAACA CCATTGTCCA CGTCTATAAT          2940

AGCTTTGGCT TTAGACCTAT TTAATGCTTC GAGACTAGCT TTAGATACTT TTAACACTCG          3000

ATTGAATTTA CTGTTATCTG CATTGACGTC AATATTGACA CGTTTCTTTT CTAATTCTGA          3060

TAATTTAGCT TCTGTTTCAG CGATATCTTT AATCAACTTT TGTTTTTGCA ACTTAACTTC          3120

TGGTGTAACT TCTTTAGAGT TTAGTTGTC TAGTTCAAAA TTCGATTCTA GTACCTTTTG           3180

TTGTAAATCT TGTATACTAG CATCTAATTT AGCTTTTACA TTTTTGTTAC TAAAGGCATC          3240

TAAAGACTTT TTAGCAACTT TGATAGTTTT TTGTAAATTT TTATCGTTAG CGTTTAATTC          3300

AACATCTTTA GTTTGATCTG CTACTCGTTT AAATCTTTGC ACAGACTTAA CCGCACTATC          3360

AATTTGCCTT TTGAATTTGG CTACACTAGC TTCAATAGTC GCTTTAATTT TATATTCCGT          3420

CACATTAACA CCTCTCTTTC TATTGCTTAT TAAATTCTGC TATAACTTTA AAGAATTCAT          3480

TATTTTGTGG TTCGTATTCA TCACGTTCGC TACTAAATCT TATATCTTTA CCTTCGTTAA          3540

GCCGTTGGAT ATTTTCTTCA TAAGGCAATA CGTCGTTTGC ATTGTTAAAA ACATATTCCT          3600

CTTTAGGTTT ATTTTCTGTC CCAACATTTT TAGTAGCTGC AGCATCACGA ATAGCAAACG          3660

CAAGTTTGTA ACGTTCGAAT TCTTGGGTTA GCATTTCATA CTCTTTCGCA TACATTCGAT          3720

AGTTATATTC TGTTAATGTC ATTTGCTCAA TAACGTTCAA ATCTGTAATA CCAAGTGTTG          3780

ACATACAAGT TATAACGATT CTGTCGTAAG TTATTAGGcT TCCGCTGGTT TTTCTTCCGT          3840

TTCCACTACT TCGACTAGGT TTCGGGTCAT AGGTCGCTTT CCCAAcTCCG TTAAAATATC          3900

CGAACGGAAT TCTTCTAGTC CGATATTTTC TGCGATTTCA TCTAATGCTT CATCAATGTT          3960

ATTAATAGTA ATTGCTTGTT TTTTTAAGTG AGATGTAGCT GCGATTAAAA cTTCGCCAAT          4020

CACAACCGGA TTTCCACTTT CTAAACCTAC AGGCAACATT GATACACCTT GACCGATAGA          4080

AGCTTGTTCA ACTTTTAAAC CTAATCGGTT ATCGATTTCT CTTAAAAATT TAAAACCAAA          4140

ACTTAATTCT AATGACTTTC CGTTAATTTC TACATTCATA ACTTAAAATC TCCATTCATA          4200

ATTAATTTAA ACAAAATAAA mArGCTTAAC GCCCTATTTT TATACCTCTC TTGGTGCAAC          4260

CGGTGGTGAA TCTACTTTAG GTTGTGGAAT TGCTGTTAAA TCTTCGCCAG TTAATGCATC          4320

TGCTTTTGTA GTGTCGTGGA ATCTGTATcC AGTCGCCTTA AGTTTCTTTG TTACAGCCTC          4380

AGGTAGTGTT GCAAATCCAC GTTGGAAACG ACCATTCACT CCATATTCAT ATTCATATTC          4440

ATCAATACCG TTAGCTTCTG CTTTTAATTC AAATTTATTG TGGAAACCTT GGAAATATTT          4500

CGCTTTAAAT TTAGCGGAAT CCCCATTTTT GCCTGGTATT CTACTTTCAA CTTCCCAAGC          4560

TTCATACAAT ACGCGATCTA CAACTGCATC TTCAATTTCA TCTGCAAAAT CGTCACCATA          4620
```

```
GTCCATTGTA TCCTCTGTAT CTGTATCAGC TTCATGTGAT AAGCCGTATT CAGTTAAAAA      4740

AAGCATTTTA GTAGCATCTA CTTTTTCGCC AGCTTTTCTA AATAAAATAA TACGATCATT      4800

ACTATTTTTC ATATTTGCCA TTCAATATTC CTCCGTTTTT TAAAATGTTT TGTAAGATAT      4860

CGTTACTGAT GTGTGTAGCA ATTCTTGATT GGTAGTATCA TCAACTAACT GTGTGATGTT      4920

AGTATCTTCT TCTTCAAAGT CATAATCGTT TGTTTTAACG CTAGGTGTTA AATCATCAAT      4980

ACATCTTTTA ACAAGTCCGT CATGATGTCC TAAATCATCG CTTACACTCC AAATATCAAT      5040

AACTAAATTC GTATCGCCAG AATAACTATC AAACGTGTAC TTACTTCTAT TTGACTCCGG      5100

CATTTTTATT ACAAAAAAAG GATACGGAAT CTCTTGTTGC ATCTCTTTAC GAGAAATAAC      5160

AGGGAATCCA TATCCTTGTA GCGTTTCATA CGCTTTATTA TAAAGTTGTA AGTTCGGTGT      5220

CATGCTTTTA TCTCCTATTC AAACAACGCT TTCAATTCTT CTACAGTTGA TTTCCTAATC      5280

ACTTCGTATA CCGGCCACAT AAAAGGTTCA GCCTCCATGT ATCGAGTACC AAATTCTAAG      5340

AAACCACTAT AAGCTGCGTG CGATGTGATA GTGTATTGCA AATCGCCAGT TTTTTTATAT      5400

CTGATATTGC GTGATaAATT ACC                                             5423
```

(2) INFORMATION FOR SEQ ID NO: 25:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6251 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```
AAACGCAGAT GTTCAATTAG AACCAGTCTA TCGTATTAAG GAAGGTATTA AACAAAAGCA       60

AATACGAGAC CAAATTAGAC AAGCGTTAAA TGATGTGACA ATTCATGAAT GGTTAACTGA      120

TGAACTAAGA GAAAAATATA AATTAGAGAC CTTGGACTTT ACTTTGAACA CATTACATCA      180

TCCTAAAAGT AAAGAGGATT TATTACGTGC TCGTAGAACC TATGCATTTA CTGAACTGTT      240

TTTATTCGAA TTACGTATGC AATGGCTAAA TAGATTAGAA AAGTCATCTG ACGAAGCAAT      300

TGAAATTGAT TATGACATAG ACCAAGTTAA ATCATTTATT GATCGTTTAC CTTTTGAACT      360

AACTGAAGCA CAGAAATCCA GTGTTAATGA AATTTTTAGA GATTTAAAAG CACCAATACG      420

TATGCATCGA TTACTTCAAG GTGATGTAGG TTCAGGAAAA ACAGTAGTTG CTGCAATTTG      480

TATGTATGCG TTAAAAACTG CTGGTTATCA ATCAGCATTG ATGGTACCAA CTGAAATTTT      540

AGCAGAGCAA CATGCTGAAA GTTTAATGGC TTTATTTGGA GATTCTATGA ACGTTGCATT      600
```

```
TACGATTGAT TGTTTAATTG GAACCCATGC TTTGATTCAA GATGATGTGA TTTTCCATAA    720

TGTTGGTTTA GTAATTACAG ATGAACAACA TCGATTTGGT GTGAATCAAC GCCAGCTTTT    780

AAGAGAAAAA GGTGCAATGA CGAATGTGTT ATTTATGACA GCAACGCCGA TACCAAGAAC    840

ACTAGCAATA TCAGTTTTTG GTGAGATGGA TGTGTCTTCA ATTAAACAAT TACCAAAAGG    900

TCGTAAACCT ATCATTACTA CTTGGGCAAA GCATGAGCAA TACGATAAAG TTTTGATGCA    960

AATGACCTCA GAGTTGAAAA AAGGTCGTCA AGCATATGTC ATTTGCCCGC TAATAGAAAG   1020

TTCTGAGCAT CTCGAAGATG TTCAAAATGT TGTCGCATTG TACGAGTCTT TACAACAGTA   1080

TTATGGTGTT TCCCGTGTAG GGTTATTGCA TGGTAAGTTA TCTGCCGATG AAAAAGATGA   1140

GGTCATGCAA AAGTTTAGTA ATCATGAGAT AAATGTTTTA GTTTCTACTA CTGTTGTTGA   1200

AGTAGGTGTT AATGTACCGA ATGCAACTTT TATGATGATT TATGATGCGG ATCGCTTTGG   1260

ATTATCAACT TTACATCAGT TACGCGGTCG TGTAGGTAGA AGTGACCAGC AAAGTTACTG   1320

TGTTTTAATT GCATCCCCTA AAACAGAAAC AGGAATTGAA AGAATGACAA TTATGACACA   1380

AACAACGGAT GGATTTGAAT TGAGTGAACG AGACTTAGAA ATGCGTGGTC CTGGAGATTT   1440

CTTTGGTGTT AAACAAAGTG GaTTGCCAGA TTTCTTAGTT GCCAATTTAG TTGAAGATTA   1500

TCGTATGTTA GAAGTTGCTC GTGATGAAGC AGCTGAACTT ATTCAATCTG GCGTATTCTT   1560

TGAAAATACG TATCAACATT TACGTCATTT TGTTGAAGAA AATTTATTAC ATCGTAGTTT   1620

TGACTAATTG CCATGCTGAT TTGTCAATTT GAGTGCAACa CTTCGTTAAT TGAGTGATAT   1680

GACACTTGAA CTATTTAAAT GTAAAGTGGT ATTTTAACAA TTTATAAATT TTCGACTAAA   1740

TAATAGCTAA ATATTACAGT TATTTGTTGA GTCGGTTAAA TAGAAAGTGT TATGATATGT   1800

GAGGAATGTT TAAGACTAGG TACTAAAAAA TGAGGGGTGA GACGTTGAAA CTAAAGAAAG   1860

ATAAACGTAG AGAAGCAATC AGACAACAAA TTGATAGCAA TCCCTTCATC ACAGACCATG   1920

AACTAAGCGA CTTATTTCAA GTGAGTATAC AAACAATTCG TTtAGaTCGC ACTTATTTAA   1980

ACATACCAGA ATTAAGGAAG CGTATTAAAT TAGTTGCTGA AAAGAATTAT GACCAAATAA   2040

GTTCTATTGA AGAACAAGAA TTTATTGGTG ATTTGATTCA AGTCAATCCa AATGTTAAAG   2100

CGCAATCAAT TTTAGATATT ACATCGGATT CTGTTTTTCA TAAAACTGGA ATTGCGCGTG   2160

GTCATGTGCT GTTTGCTCAG GCAAATTCGT TATGTGTTGC GCTAATTAAG CAACCAACAG   2220

TTTTAACTCA TGAGAGTAGC ATTCAATTTA TTGAAAAAGT AAAATTAAAT GATACGGTAA   2280

GAGCAGAAGC ACGAGTTGTA AATCAAACTG CAAAACATTA TTACGTCGAA GTAAAGTCAT   2340

ATGTTAAACA TACATTAGTT TTCAAAGGAA ATTTTAAAAT GTTTTATGAT AAGCGAGGAT   2400
```

```
TTAGAAGCCG TACAAAAGGC TGTTGAAGAC TTTAAAGATC TAGAAATTAT ACTTTTCGGT    2520

GACGAAAAAA AGTATAATCT GAACCATGAA CGAATCGAAT TTAGACATTG TTCTGAAAAG    2580

ATTGAAATGG AAGATGAGCC TGTTAGAGCG ATTAAACGTA AAAAAGATAG CTCAATGGTA    2640

AAAATGGCTG AAGCTGTGAA ATCTGGTGAA GCAGATGGAT GTGTGTCAGC AGGTAATACT    2700

GGTGCTTTAA TGTCAGCTGG TTTATTCATT GTTGGACGTA TTAAAGGTGT AGCTAGACCG    2760

GCTTTAGTAG TAACATTGCC AACGATTGAT GGAAAAGGTT TTGTCTTTTT AGACGTTGGT    2820

GCAAATGCTG ATGCTAAACC TGAACACTTA TTACAGTATG CGCAACTAGG GGATATTTAT    2880

GCTCAAAAAA TTAGAGGTAT TGATAATCCG AAAATCTCAT TATTAAATAT AGGAACCGAG    2940

CCAGCTAAAG GTAATAGTTT AACGAAAAAA TCATATGAGT TATTAAATCA TGATCATTCA    3000

TTGAATTTTG TTGGGAATAT TGAAGCGAAG ACATTAATGG ATGGCGATAC AGATGTTGTA    3060

GTTACCGATG GCTATACTGG GAACATGGTC CTTAAAAATT TAGAAGGTAC TGCAAAATCA    3120

ATCGGTAAAA TGTTAAAAGA TACGATTATG AGTAGTACTA AAAATAAATT AGCAGGTGCA    3180

ATATTGAAGA AAGATTTAGC TGAATTCGCT AAAAAGATGG ATTACTCAGA ATACGGTGGT    3240

TCCGTATTAT TAGGATTGGA AGGTACTGTA GTTAAAGCAC ACGGTAGTTC AAATGCTAAA    3300

GCTTTTTATT CTGCAATTAG ACAAGCGAAA ATCGCAGGAG AACAAAATAT TGTACAAACA    3360

ATGAAAGAGA CTGTAGGTGA AtCAAATGaG TaAAACAGCA ATTATTTTTC CGGGACAAGG    3420

TGCCCAAAAA GTTGGTATGG CGCAAGATTT GTTTAACAAC AATGATCAAG CAACTGAAAT    3480

TTTAACTTCA GCAGCGAACA CATTAGACTT TGATATTTTA GAGACAATGT TTACTGATGA    3540

AGAAGGTAAA TTGGGTGAAA CTGAAAACAC ACAACCAGCT TTaTTGaCGC aTAGTTCGGC    3600

ATTATTAGCA GCGCTAAAAA ATTTGAATCC TGATTTTACT ATGGGGCATA GTTTAGGTGA    3660

ATATTCAAGT TTAGTTGCAG CTGACGTATT ATCATTTGAA GATGCAGTTA AAATTGTTAG    3720

AAAACGTGGT CAATTAATGG CGCAAGCATT TCCTACTGGT GTAGGAAGCA TGGCTGCAGT    3780

ATTGGGATTA GATTTTGATA AAGTCGATGA AATTTGTAAG TCATTATCAT CTGATGACAA    3840

AATAATTGAA CCAGCAAACA TTAATTGCCC AGGTCAAATT GTTGTTTCAG GTCACAAAGC    3900

TTTAATTGAT GAGCTAGTAG AAAAAGGTAA ATCATTAGGT GCAAAACGTG TCATGCCTTT    3960

AGCAGTATCT GGACCATTCC ATTCATCGCT AATGAAAGTG ATTGAAGAAG ATTTTTCAAG    4020

TTACATTAAT CAATTTGAAT GGCGTGATGC TAAGTTTCCT GTAGTTCAAA ATGTAAATGC    4080

GCAAGGTGAA ACTGACAAAG AAGTAATTAA ATCTAATATG GTCAAGCAAT TATATTCACC    4140

AGTACAATTC ATTAACTCAA CAGAATGGCT AATAGACCAA GGTGTTGATC ATTTTATTGA    4200
```

```
AACATCAATT CAAACTTTAG AAGATGTGAA AGGATGGAAT GAAAATGACT AAGAGTGCTT        4320

TAGTAACAGG TGCATCAAGA GGAATTGGAC GTAGTATTGC GTTACAATTA GCAGAAGAAG        4380

GATATAATGT AGCAGTAAAC TATGCAGGCA GCAAAGAGAA AGCTGAAGcA GTAGTCGAAG        4440

AAATCAAAGC TAAAGGTGTT GACAGTTTTG CGATTCAAGC AAATGTTGCC GATGCTGATG        4500

AAGTTAAAGC AATGATTAAA GAAGTAGTTA GCCAATTTGG TTCTTTAGAT GTTTTAGTAA        4560

ATAATGCAGG TATTACTCGC GATAATTTAT TAATGCGTAT GAAAGAACAA GAGTGGGATG        4620

ATGTTATTGA CACAAACTTA AAAGGTGTAT TTAACTGTAT CCAAAAAGCA ACACCACAAA        4680

TGTTAAGACA ACGTAGTGGT GCTATCATCA ATTTATCAAG TGTTGTTGGA GCAGTAGGTA        4740

ATCCGGGACA AGCAAACTAT GTTGCAACAA AAGCAGGTGT TATTGGTTTA ACTAAATCTG        4800

CGGCGCGTGA ATTAGCATCT CGTGGTATCA CTGTAAATGC AGTTGCACCT GGTTTTATTG        4860

TTTCTGATAT GACAGATGCT TTAAGTGATG AGCTTAAAGA ACAAATGTTG ACTCAAATTC        4920

CGTTAGCACG TTTTGGTCAA GACACAGATA TTGCTAATAC AGTAGCGTTC TTAGCATCAG        4980

ACAAAGCAAA ATATATTACA GGTCAAACAA TCCATGTAAA TGGTGGAATG TACATGTAAT        5040

ATATTTGAGC TAAAGCTCAT TGACGCAGTG GTTGACTGGT CATCCAATGG AGAATTGTCT        5100

GACCTAGTCA ACTTTGCGGG GGAAATTCTA AGCAACCTAG ATAAGGTTCC AGAATTTCTC        5160

CCTAAGAAAC ACTAATCAAT aAATTGwTAA GTGTTTCTAA AATTTCTACT TGTTTTTTAG        5220

AATTTAAAAT GGGAAAATAT AGTAGTCTAT GTATAGGCAT TTTTAAAGGA GGTGAATCGA        5280

CGTGGAAAAT TTCGATAAAG TAAAAGATAT CATCGTTGAC CgTTTAGGTG TAGACGCTGA        5340

TAAAGTAACT GAAGATGCAT CTTTCAAAGA TGATTTAGGC GCTGACTCAC TTGATATCGC        5400

TGAATTAGTA ATGGAATTAG AAGACGAGTT TGGTACTGAA ATTCCTGATG AAGAnGCTGA        5460

AAAAaTCAAC ACTGTTGGTG ATGCTGTTAA ATTTATTAAC AGTCTTGAAA AATAATAAAT        5520

CTTACATCTG GGTCGTCAGT ATTGTCGACT CAGTTTTTTT CTTTAATTAT CAATAGTTTT        5580

AACGTAAAAT TAAAGATGAT TCAAGAGCAA CACATAAAGG AGATAAAATA ATGTCTAAAC        5640

AAAAGAAAAG TGAGATAGTT AATCGTTTTA GAAAGCGCTT TGATACTAAA ATGACAGAGT        5700

TAGGCTTTAC TTATCAAAAT ATTGATTTAT ACCAACAAGC ATTTTCGCAT TCGAGTTTTA        5760

TTAATGATTT TAATATGAAT CGTTTAGACC ATAATGAGCG TTTAGAGTTT TTGGGTGATG        5820

CGGTATTAGA ATTGACGGTT TCACGATATT TATTTGATAa ACATCCCAAC TTGCCAGAAG        5880

GGAATTTAAC AAAAATGCGT GCCaCTATTG TATGTGAGCC CtCACTkGTA ATATTTGCGA        5940

ATAAAATTGG ATTGAACGAA ATGATTTTAC TTGGTAAAGG TGAAGAGAAA ACAGGGGGAC        6000
```

ATCAAGGACT AGATATAGTT TGGAAATTTG CTGAGAAAGT CATTTTCCCA CATGTAGAAC 6120

AAAATGAGTT ATTAGGCGTG GTAGATTTTA AAACACAATT CCAAGAATAT GTGCACCAGC 6180

AAAATAAAGG TGATGTAACC TATAATTTAA TAAAAGAAGA GGGACCGGCA CATCATCGTC 6240

TATTCACTTC A 6251

(2) INFORMATION FOR SEQ ID NO: 26:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4920 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

ACCTACTGAA GTTGCTAATT TTTTGGAGCA ACTAAGCACT GAAATTGAAC GTCTTAAAGA 60

AGATAAAAAA CAACTTGAAA AAGTAATCGA AGAGAGaGAT ACTAATATTA AGTCTTATCA 120

AGACGTGGcA TCAATCTGTA AGTGaTGCTT TGATACAAGC TCAAAAAGCT GGTGAAGAAA 180

CTAAGCAAGC TGCAGAGAAA CAAGCTGAAG CGATTATAGC TAAGGCAGAA GCGCAAGcTA 240

ATCAAATGGT TGGTGACGCG GTAGAAAAAG CACGCCGTTT AGCATTCCAG ACTGAAGATA 300

TGAAACGTCA ATCAAAAGTA TTTAGATCGC GTTTCCGTAT GTTAGTTGAA GCGCAATTAG 360

ACTTATTAAA AAACGAAGAT TGGGATTACT TGTTGAATTA TGATTTAGAC GCTGAACAAG 420

TGACGCTTGA AAATATTCAT CATTTGCATG AAAATGATTT AAAGCCAGAT GAAGTTGCAG 480

CAAATGCACA AAATAATGCA TCAAATACAC CAGACAATAA TCAACAATCC AATGATTCAG 540

AAACAACTAA GAAGTAAGAA TTAAATAAAG ACAGACGCGT AATATACATT TAACTTTTCA 600

CAGCGAATTA GGTAATGGTG AGAGCCTAGT AAAAGCATGT ATGTTATATC ACTGGCTTTT 660

TAATATTTAA ATAATGTAAT GAGAGAACTC TAAGTTGAGT TAATAAGGGT GGTACCGCGA 720

GCAATCGTCC CTTTTAATTT AACTTAGAGT TTTTTAAATT TTTAAGGAGT GAAAAAAATG 780

GATTACAAAG AAACGTTATT AATGCCTAAA ACAGATTTCC CAATGCGAGG TGGTTTACCA 840

AACAAGGAAC CGCAAATTCA AGAAAAATGG GATGCAGAAG ATCAATACCA TAAAGCGTTA 900

GAAAAAAATA AAGGTAACGA AACATTCATT TTACATGATG GCCCACCATA CGCGAATGGT 960

AACTTACATA TGGGACATGC CTTGAACAAA ATTTTAAAAG ACTTTATTGT ACGTTATAAA 1020

ACTATGCAAG GGTTCTATGC ACCATACGTA CCAGGTTGGG ATACACATGG TTTACCAATT 1080

GAACAAGCAT TAACGAAAAA AGGTGTTGAC CGAAAGAAAA TGTCAACAGC TGAATTCCGT 1140

```
TTAGGTGTTC GTGGTGACTT TAATGATCCA TATATTACAT TAAAACCTGA ATACGAAGCT    1260

GCACAAATTC GTATTTTTGG AGAAATGGCA GATAAAGGTT TAATTTATAA AGGTAAAAAG    1320

CCAGTTTATT GGTCTCCTTC AAGTGAGTCT TCATTAGCAG AAGCAGAAAT TGAATATCAC    1380

GATAAACGTT CAGCATCAAT TTACGTTGCA TTTGACGTTA AAGATGACAA AGGTGTCGTT    1440

GATGCAGATG CTAAATTTAT TATCTGGACA ACAACGCCAT GGACAATTCC ATCAAATGTT    1500

GCGATTACCG TTCATCCTGA ATTAAAATAT GGTCAATACA ATGTAAATGG cGAAAAATAT    1560

ATTATTGCAG AAGCCTTGTC TGACGCTGTA GCAGAAGCAC TGGaTTGGGA TAAAGCATCA    1620

ATCAAATTAG AAAAAGAATA CACAGGTAAA GAATTAGAGT ATGTTGTAGC ACAACATCCA    1680

TTCTTAGACA GAGAATCGTT AGTGATTAAT GGTGATCATG TTACTACAGA TGCTGGTACA    1740

GgTTGTGTAC ATACAGCACC AGGTCACGGG GAAGATGACT ATATTGTTGG TCAAAAATAT    1800

GAATTGCCAG TAATTAGTCC AATCGATGAT AAAGGTGTAT TTACTGAAGA AGGCGGCCAA    1860

TTTGAAGGGA TGTTCTATGA TAAAGCTAAT AAAGCCGTTA CTGATTTATT AACAGAAAAA    1920

GGTGCACTAT TAAAAATTAGA CTTTATTACA CATAGCTATC CACACGACTG GAGAACAAAA    1980

AAACCTGTAA TCTTCCGTGC TACACCACAA TGGTTTGCCT CAATCAGTAA AGTAAGACAA    2040

GATATTTTAG ATGCAATCGA AAATACAAAC TTCAAAGTAA ATTGGGGTAA AACACGTATT    2100

TACAATATGG TTCGTGACCG TGGCGAATGG GTTATTTCTC GTCAACGTGT GTGGGGTGTA    2160

CCGTTACCAG TATTTTATGC TGAAAATGGC GAAATTATCA TGACGAAAGA AACAGTGAAT    2220

CATGTTGCTG ATTTATTTGC AGAACACGGT TCAAATATTT GGTTGAAAG AGAAGCGAAA    2280

GACTTACTAC CAGAAGGATT TACACATCCA GGCAGCCCTA ACGGTACATT TACTAAAGAA    2340

ACAGACATTA TGGACGTTTG GTTTGATTCT GGTTCATCAC ACCGTGGCGT GTTGGAAACA    2400

AGACCGGAAT TAAGTTTCCC AGCGGATATG TATTTAGAAG GTAGTGACCA ATATCGTGGT    2460

TGGTTCAACT CTTCTATCAC AACTTCAGTT GCTACAAGAG GAGTATCACC TTATAAATTC    2520

TTACTTTCTC ATGGTTTTGT TATGGACGGT GAAGGTAAGA AAATGAGTAA ATCTTTAGGT    2580

AATGTGATTG TACCTGACCA AGTGGTTAAA CAAAAAGGTG CTGATATTGC GAGACTTTGG    2640

GTAAGTAGTA CGGACTATTT AGCTGATGTT AGAATTTCTG ATGAAATTTT AAAACAAACA    2700

TCTGATGTTT ATCGTAAAAT CAGAAATACA TTAAGATTTA TGTTAGGTAA CATTAACGAT    2760

TTCAATCCTG ACACAGATAG CATTCCTGAA TCAGAGTTAT TAGAAGTGGA TCGTTACTTG    2820

CTAAATCGTT TACGTGAATT TACTGCAAGT ACGATTAACA ACTATGAAAA CTTTGACTAC    2880

TTAAATATTT ATCAAGAAGT TCAAAACTTT ATCAATGTTG AGTTAAGTAA TTTCTATTTG    2940
```

```
CAAACAGTGT TATATCAAAT TTTAGTTGAT ATGACGAAGT TGTTAGCACC AATCTTAGTG        3060

CATACAGCTG AAGAAGTTTG GTCTCATACA CCACATGTTA AAGAAGAAAG TGTTCACTTA        3120

GCAGACATGC CTAAAGTTGT AGAAGTAGAT CAAGCTTTAT TGGATAAATG GCGTACATTT        3180

ATGAATTTAC GTGATGATGT GAACCGTGCA TTAGAAACTG CTCGTAATGA AAAAGTTATT        3240

GGTAAATCAT TAGAAGCTAA AGTTACGATT GCTAGTAACG ATAAATTTAA TGCATCTGAA        3300

TTCTTAACTT CATTTGATGC ATTACATCAA TTATTTATCG TGTCACAAGT TAAAGTTGTA        3360

GATAAGTTAG ACGATCAGGC AACAGCTTAT GAACATGGTG ATATTGTCAT CGAACATGCA        3420

GATGGTGAAA AATGTGAAAG ATGTTGGAAC TATTCAGAGG ATCTTGGTGC TGTTGATGAA        3480

TTGACGCATC TATGTCCACG ATGCCAACAA GTTGTAAAAT CACTTGTATA ATTGAAATTG        3540

TATAAAGTAC TCATACAGAT GATATAAATT AAAGCTCTCT TCATAATCAT GTTGTAGTTT        3600

TTGTTGACAT GATGAAGAGA GTTTTTTTGT GAATAAAAAA ATGACCAAGT TACCGGTCAT        3660

ATATGTAAAA AATGTGCGAT TTACTAAAAT AAAAATTATT CAGGAATGGT ACAAATTCTC        3720

TGAGGCATAT AAATGCGTTA TAGTTGCTAT TCTCAATTAT GTTCGCGATA ATTTTAAGTA        3780

AAAGTAAGCA CAGATATTGA ATTTGATAGG AGTTAATTGA ATGTATCATA ACAGTAACGC        3840

AAACTTTGTC AATGGTATCA CTTTAAATGT GAGAGATAAG AATGAATTAA AGCCATTTTA        3900

TGAGGACATA TTAGGATTAA ATATTATAAA TGAGACATTA ACATCGATAC AATATGAAGT        3960

AGGTCAAAAT AATCATGTCA TTACACTTGT TGAATTACAA AATGGACGTG AACCTTTAAT        4020

GTCCGAAGCG GGACTGTTTC ATATCGCAAT TAAACTACCT CAAATTAGTG ATTTAGCTAA        4080

TTTACTAATT CATTTAAGCG AATATGATAT TCCAGTTAAC GGAGGTATAC AGCCTGCTTC        4140

GTTATCATTA TTTTTTGAAG ACCCGGAAGG AAACGGTTTT AAATTTTATG TTGATAAAGA        4200

CGAAGCGCAA TGGACGAGGC AAAATAATTT AGTAAAAATT GATATTAGAC CATTAAATGT        4260

ACCGAGATTA GTGAGTCATG CAACAAAATT GTTATGGTTA GGTATTCCAG ATGACGCTAT        4320

TATAGGTGCA TTGCATATTA AGACAATTCA TTTATCAGAG GTAAAAGAGT ACTACCTCGA        4380

TTATTTTGGA TTAGAGCAAT CGGCATATAT GGATGATTAT TCAATATTTT TAGCATCGAA        4440

TGGCTATTAT CAACATTTGG CCATGAATGA TTGGGTATCA GCAACGAAAC GTGTAGAAAA        4500

TTTTGATACG TATGGATTAG CAATTGTTGA CTTTCATTAT CCTGAAACAA CACATTTAAA        4560

TTTACAAGGT CCGGATGGTA TCTATTATCG CTTTAATCAT ATCGAAGTTG AAGATTAGTA        4620

TATACTTTGA ATGGACGAAC CATATAATGA ATCGTTTTTA ATGATCTTTT TATACAAGTT        4680

ATGAAGGAGG CTGGGACATT AAGTTCTTAG GCAATGTAAA AAGCTGATTT CTATTAATTA        4740
```

302

TTTTCCTTAT ATTAATTGCC ATTAATACAA AACCTAGCTC TCGTTTAACT TTATTTATTC   4860

CTCGAACTGA CATTCGnGTG AACTCAAAAT nGCCTACTTn CTTAAATTAC CAATATCTAT   4920

(2) INFORMATION FOR SEQ ID NO: 27:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 626 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TGGATTGCCA TTACATGGAC AAGATTTAAC TGAATCAATT ACACCATATG AAGGTGGTAT   60

CGCTTTTGCA AGTAAACCAT TAATTGATGC TGATTTTATT GGTAAATCTG TATTAAAAGA   120

TCAAAAAGAA AATGGTGCAC CAAGAAGAAC AGTGGGATTA GAATTACTTG AAAAAGGAAT   180

TGCAAGAACT GGTTATGAAG TTATGGATTT AGATGGAAAT ATTATTGGAG AAGTAACTTC   240

AGGAACACAG TCTCCATCAT CAGGAAAATC AATTGCACTT GCAATGATAA AAAGAGATGA   300

GTTTGAAATG GGTAGAGAGT TGCTTGTTCA AGTTCGTAAG CGTCAATTAA AAGCGAAAAT   360

TGTTAAGAAA AATCAAATTG ATAAATAATT AAAAAGGGGT GTGCATTGTG AGTCATCGTT   420

ATATACCTTT AACTGAAAAA GACAAGCAAG AAATGTTACA AACAATTGGT GCAAAATCTA   480

TAGGAGAATT ATTCGGTGAT GTACCAAGTG ACATTTTATT AAATAGAGAT TTAAATATTG   540

CTGAAGGCGA ACGGAGAACA ACGTTACTTA GAAGATTnAA TCGCATTGCA AGCAAGAGTA   600

TCACTAGAGG AACGCGTACA TCGTTT   626

(2) INFORMATION FOR SEQ ID NO: 28:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1126 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

nGGAAGTGGT GTATATATTT GTAATGAGTG TATTGAATTA TGCTCAGAAA TCGTCGAAGA   60

AGAATTAGCT CAAAACACTT CTGAAGCGAT GACAGAATTA CCTACTCCTA AAGAAATTAT   120

GGATCATTTA AACGAATATG TTATTGGTCA AGAAAAAGCT AAAAAATCTT TAGCTGTAGC   180

TGTTTATAAC CACTATAAGC GTATTCAACA ATTAGGACCA AAAGAAGATG ATGTTGAATT   240

```
AACCTTAGCC AAGACGTTGA ATGTACCATT TGCAATTGCA GATGCGACAA GTTTAACTGA    360

AGCTGGTTAT GTAGGCGATG ATGTTGAAAA TATCTTGTTG AGATTAATTC AAGCAGCTGA    420

CTTTGACATT GATAAAGCCG AAAAAGGTAT TATTTATGTA GATGAAATTG ATAAAATTGC    480

ACGTAAATCT GAAAACACAT CTATAACACG TGACGTTTCA GGTGAAGGTG TTCAACAAGC    540

ATTGCTTAAA ATCTTAGAAG GTACGACTGC AAGTGTTCCG CCACAAGGTG GACGCAAACA    600

TCCAAACCAA GAAATGATTC AAATTGATAC AACAAATATC TTATTTATTC TTGGTGGTGC    660

CTTTGATGGT ATTGAAGAAG TGATTAAGCG CCGTCTTGGT GAAAAAGTTA TTGGTTTCTC    720

AAGCAATGAA GCTGATAAAT ATGACGAACA AGCATTATTA GCACAATTC GCCCAGAAGA    780

TTTGCAAGCC TATGGTTTGA TTCCTGAATT TATCGGACGT GTGCCAATTG TAGCTAATTT    840

AGAAACATTA GATGTAACTG CGTTGAAAAA CATCTTAACG CAACCTAAAA ATGCACTTGT    900

GAAACAATAT ACTAAAATGC TGGAATTAGA TGATGTGGAT TTAGAGTTCA CTGAAGAAGC    960

TTTATCAGCA ATTAGTGAAA AAGCAATTGA AAGAAAAACA GGTGCGCGTG GTTTACGTTC   1020

AATCATAGAA GAATCGTTAA TCGATATTAT GTTTGATGTG CCTTCTAACG AAAATGTAAC   1080

GAaGGTAGTT ATTACAGCAC AAACmATTAA TGrAGaACTG AACCAG                  1126
```

(2) INFORMATION FOR SEQ ID NO: 29:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 4392 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

```
ATTGACTTCT TAGCAATnAA TaTGAGTGAA GAACGTACTG TTGAAGTACC AGTTCAATTA     60

GTTGGTGAAG CAGTAGGCGC TAAAGAAGGC GGCGTAGTTG AACAACCATT ATTCAACTTA    120

GAAGTAACTG CTACTCCAGA CAATATTCCA GAAGCAATCG AAGTAGACAT TACTGAATTA    180

AACATTAACG ACAGCTTAAC TGTTGCTGAT GTTAAAGTAA CTGGCGACTT CAAAATCGAA    240

AACGATTCAG CTGAATCAGT AGTAACAGTA GTTGCTCCAA CTGAAGAACC AACTGAAGAA    300

GAAATCGAAG CTATGGAAGG CGAACAACAA ACTGAAGAAC CAGAAGTTGT TGGCGAAAGC    360

AAAGAAGACG AAGAAAAAAC TGAAGAGTAA TTTTAATCTG TTACATTAAA GTTTTTATAC    420

TTTGTTTAAC AAGCACTGTG CTTATTTTAA TATAAGCATG GTGCTTTTTG TGTTATTATA    480

AAGCTTAATT AAACTTTATT ACTTTGTACT AAAGTTTAAT TAATTTTAGT GAGTAAAAGA    540
```

```
CTTACTAAGC TAAAGAATAA TGATAATTGA TGGCAATGGC GGAAAATGGA TGTTGTCATT    660

ATAATAATAA ATGAAACAAT TATGTTGGAG GTAAACACGC ATGAAATGTA TTGTAGGTCT    720

AGGTAATATA GGTAAACGTT TTGAACTTAC AAGACATAAT ATCGGCTTTG AAGTCGTTGA    780

TTATATTTTA GAGAAAAATA ATTTTTCATT AGATAAACAA AAGTTTAAAG GTGCATATAC    840

AATTGAACGA ATGAACGGCG ATAAAGTGTT ATTTATCGAA CCAATGACAA TGATGAATTT    900

GTCAGGTGAA GCaGTTGCAC CGATTATGGA TTATTACAAT GTTAATCCAG AAGATTTAAT    960

TGTCTTATAT GATGATTTAG ATTTAGAACA AGGACAAGTT CGCTTAAGAC AAAAAGGAAG   1020

TGCGGGCGGT CACAATGGTA TGAAATCAAT TATTAAAATG CTTGGTACAG ACCAATTTAA   1080

ACGTATTCGT ATTGGTGTGG GAAGACCAAC GAATGGTATG ACGGTACCTG ATTATGTTTT   1140

ACAACGCTTT TCAAATGATG AAATGGTAAC GATGGAAAAA GTTATCGAAC ACGCAGCACG   1200

CGCAATTGAA AAGTTTGTTG AAACATCACG ATTTGACCAT GTTATGAATG AATTTAATGG   1260

TGAAGTGAAA TAATGACAAT ATTGACAACG CTTATAAAAG AAGATAATCA TTTTCAAGAC   1320

CTTAATCAGG TATTTGGACA AGCAAACACA CTAGTAACTG GTCTTTCCCC GTCAGCTAAA   1380

GTGACGATGA TTGCTGAAAA ATATGCACAA AGTAATCAAC AGTTATTATT AATTACCAAT   1440

AATTTATACC AAGCAGATAA ATTAGAAACA GATTTACTTC AATTTATAGA TGCTGAAGAA   1500

TTGTATAAGT ATCCTGTGCA AGATATTATG ACCGAAGAGT TTTCAACACA AAGCCCTCAA   1560

CTGATGAGTG AACGTATTAG AACTTTAACT GCGTTAGCTC AAGGTAAGAA AGGGTTATTT   1620

ATCGTTCCTT TAAATGGTTT GAAAAAGTGG TTAACTCCTG TTGAAATGTG GCAAAATCAC   1680

CAAATGACAT TGCGTGTTGG TGAGGATATC GATGTGGACC AATTTCTTAA CAAATTAGTT   1740

AATATGGGGT ACAAACGGGA ATCCGTGGTA TCGCATATTG GTGAATTCTC ATTGCGAGGA   1800

GGTATTATCG ATATCTTTCC GCTAATTGGG GAACcAATCA GAATTGAGCT ATTTGATACC   1860

GAAATTGATT CTATTCGGGA TTTTGATGTT GAAACGCAGC GTTCCAAAGA TAATGTTGAA   1920

GAAGTCGATA TCACAACTGC AAGTGATTAT ATCATTACTG AAGAAGTGAT CAGCCATCTT   1980

AAAGAAGAGT TAAAAACTGC ATATGAAAAT ACAAGACCCA AAATAGATAA ATCAGTGCGC   2040

AATGATTTGA AAGAAACGTA TGAAAGCTTT AAATTATTCG AAAGTACATA CTTTGATCAT   2100

CAAATACTAC GTCGCTTAGT AGCGTTTATG TATGAAACAC CTTCGACAAT TATTGAGTAT   2160

TTCCAAAAAG ATGCAATCAT TGCAGTTGAT GAATTTAATC GTATTAAAGA AACTGAAGAA   2220

AGTTTAACAG TAGAGTCTGA TTCGTTTATT AGCAATATTA TTGAAAGTGG TAATGGATTT   2280

ATAGGACAAA GTTTTATAAA ATATGATGAT TTTGAAACAT TGATTGAAGG CTATCCTGTC   2340
```

```
TCATGTAAAC CTGTCCAACA ATTTTATGGG CAATATGACA TTATGCGTTC TGAATTTCAA    2460

CGATATGTTA ATCAAAACTA TCATATCGTG GTTTTGGTCG AAACCGAAAC TAAAGTTGAA    2520

CGTATGCAAG CGATGTTAAG TGAAtGCAT ATTCCATCAA TAACAAAATT GCATCGCTCA     2580

ATGTCATCGG GGCAAGCAGT GATTATTGAA GGCAGTTTAT CTGAAGGATT TGAACTACCT    2640

GATATGGGAT TAGTTGTCAT TACTGAGCGT GAgcTTTTTA AATCAAAACA GAAAAAGCAA    2700

CGAAAACGTA CGAAAGCTAT CTCAAATGCT GAAAAAATTA AGTCTTACCA AGATTTAAAT    2760

GTGGGAGATT ATATTGTTCA TGTGCATCAT GGTGTTGGTA GATATTTAGG TGTTGAGACG    2820

CTCGAAGTGG GGCAAACGCA TCGTGATTAT ATTAAATTGC AATATAAAGG TACGGATCAA    2880

CTATTTGTTC CAGTAGATCA AATGGATCAA GTTCAAAAAT ATGTAGCTTC GGAAGATAAG    2940

ACGCCAAAAT TAAATAAACT CGGTGGCAGT GAATGGAAAA AAACAAAAGC TAAAGTTCAA    3000

CAAAGTGTTG AAGATATTGC TGAAGAGTTG ATTGATTTAT ATAAAGAAAG AGAAATGGCA    3060

GAAGGTTATC AATATGGGGA AGACACAGCT GAGCAAACAA CATTTGAATT AGATTTTCCA    3120

TATGAACTTA CGCCTGACCA AGCTAAATCT ATCGATGAAA TTAAAGATGA CATGCAAAAA    3180

TCGCGTCCAA TGGATCGCTT GCTATGTGGT GATGTTGGTT ATGGTAAAAC TGAAGTTGCA    3240

GTGAGAGCAG CATTCAAAGC TGTAATGGAA GGAAAGCAGG TTGCATTTTT AGTTCCTACA    3300

ACTATTTTAG CTCAGCAACA TTATGAGACG TTAATTGAGC GTATGCAAGA TTTTCCTGTT    3360

GAAATTCAAT TAATGAGTCG TTTTAGAACG CCTAAAGAGA TAAAACAAAC TAAGGAAGGA    3420

CTTAAAACTG GATTTGTTGA CATAGTTGTT GGTACACACA AATTACTTAG TAAAGATATA    3480

CAGTATAAAG ATTTAGGGCT GTTGATTGTA GATGAAGAAC AACGATTTGG TGTACGCCAT    3540

AAAGAGCGTA TTAAAACATT AAAACATAAT GTAGATGTAC TAACATTGAC TGCAACCCCA    3600

ATAGCTAGAA CATTGCATAT GAGTATGCTA GGTGTGCGGG ATTTGTCAGT GATTGAAACG    3660

CCGCCAGAAA ATCGTTTCCC AGTTCAAACA TATGTATTAG AACAGAACAT GAGTTTTATC    3720

AAAGAAGCTT TAGAAAGAGA ACTATCCCGT GATGGCCAAG TGTTTTATCT TTATAATAAA    3780

GTGCAATCCA TTTATGaAAA ACGAGAACAA CTCCAGATGT TAATGCCAGA TGCTAACATT    3840

GCAGTTGCTC ATGGACAAAT GACAGAGCGC GATTTAGAAG AAACGATGTT AAGTTTTATC    3900

AATAATgAAT ATGATATTTT AGTAACGACG ACGATTATTG AAACAGGTGT CGATGTCCCA    3960

AATGCAAATA CTTTGATCAT TGAAGATGCA GATCGCTTTG GATTGAGTCA GTTGTATCAA    4020

TTAAGAGGTC GTGTTGGTCG TTCAAGTCGT ATTGGTTATG CATACTTCTT ACATCCAGCA    4080

AATAAGGTAC TAACTGAGAC TGCAGAAGAT CGATTACAAG CGATTAAAGA ATTTACGGAG    4140
```

306

```
TTAGGTAAAC AACAGCACGG CTTTATTGAT ACAGTTGGAT TTGATTTGTA CAGTCAAATG        4260

TTAGAAGAAG CTGTAAATGA AAAACGTGGT ATTAAGGAAC CAGAATCTGA GGTGCCAGAA        4320

GTCGAAGTTG ATTTAAACTT GGATGCATAT TTGCCAACAG AATATATTGC AAATGAACAA        4380

GCTAAAATTG AA                                                           4392
```

(2) INFORMATION FOR SEQ ID NO: 30:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 729 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

```
TTTCTTTTGA ATCTATATCG AGGTGGTTGG TAGGTTCATC TAAAATAAGT ACATTGTCAC          60

GTTGCAACAT AAGTAGTGCT AGTTGTAAAC GTGCTTTTTC ACCACCAGAT AAATCATTAA         120

TTATCTTTTT AACATCGTCT TGTACAAATA AGAAACGTCC AAGAACTGCT CGAATATCTT         180

TTTCATTCAT TAACGGATAT TGATCCCACA CATAATCTAA AATCGTTTTA CTAGATTTAA         240

ATTCTGCTTG CTTTTGATCA TAATAACCAA TTTGTAAATT TGCGCCGAAA GTAATATCGC         300

CATTAAGCGC TTTTTGTTGA TTAGCAATAG TTTTAATTAA GGTCGATTTT CCAATACCAT         360

TTGGCCCAAT GATTGCTATA TGATCGCCTT TAGAGACCTC TATACTCATA GGTTTGGTAA         420

TTGCAGTTTG ATAACCGATT TCTAAATTTT TTACATGCAT GACGTCATTA CCTGTATTCC         480

GGTCAAAGCC AAATTGAATA TTTGCACTTT TGGCATCTAA CATTGGTTTA TCAATGCGTT         540

CCATTTTTTC TAAAATCTTA CGTCTACTTT TTGCCATTCC ACTTGTTGAA GCACGGGTAA         600

TATTTTTCTC AACAAAAGTT TCTAATCGTT TTATTTCTGC TTGTTGACTT TCATATTCTT         660

GCATTCGTTT TTGATAATAT AAATCCCGTT GCTGTATAAA TTCCTCGTAA TTACCAACAT         720

AGCGTTTGA                                                                729
```

(2) INFORMATION FOR SEQ ID NO: 31:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 13856 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TGATGTTTCG ATACATTTGT TGCACCTTGT GGATATACTT TAAAGGTTGT GTCGTATGTT   120

TCCTTACTAT CTTTAGCTTC AGATTCCTGT GATTCAACCG TTTTATATTT TTCAAGTGCA   180

TGTCCTTCAA TATCAACTCG TGGAATAATG CGATTCAACC ATGCTGGTAA ATACCACGAA   240

CCTTTtCCAA ACAATTTCGt TAATGCAGGA ATTAACATCA TtCTGACTAC GAAGGCATCA   300

AAGAGTACAC CAAACGCTAA TGCCATACCC ATTGATTTAA TCATGACATC TTCTTGGAAT   360

ACAAACGCAA AGAAGACACT AAACATAATT AATGCAGCTG CTACAATAAC AGGACCGCTT   420

TCTTTCAATC CTACTTTGAT AGAATAATCA TTATCCCCTG TTTTACTATTm yyCTTCATGr   480

ATTCGCGACA TAAGGAAGAC TTCATAATCC ATCGCTAATC CAAATAAGAT ACCTATAGTA   540

ATAACCGGTA AAAATGCTAG CATTGGTCCT GTCGTTTCAA TACCAAACAG ACCTTTCATA   600

AAACCATCTT GCATTACTAA TGTTGTAAAT CCTAATGTTG CCATTAATGA CAAGACGAAT   660

CCTAAAACTG CTTTTAATGG TATTAGAATT GAACGGAAGA CAATCATTAA TAAGAAAAAT   720

GCTAATACAA CAATGACTGA GGCAAATAAA GGTATCGCCT CATTTAACTT TTTAGACATA   780

TCAATATTAA TGACACTTTG TCCCGAAATC TCCGTTTTGA ACCCATATTT ATCTTGTGCA   840

TCTTTATGAT AATCTCGTAA ATCATGCACT AAATCATTTG TACTCTCTGC ATTAGGCCCT   900

TGCTTAGGTA TCACGACCAT CAAAGCGTAA TCATTATCTT TACTCATTTG TGGTGGCGTA   960

ACGATATCTA CATTTTTCTT ATCTTTAATA TCTTTATATA CAGACTGTAA ATCTTGTTGT   1020

AATCCTTGTG GATCATCCTT TTTATCTTTC ACATTTATCA ACATCGGTAT TTGGCCATTA   1080

AATCCTTCAC CAAATTTATC CGAGATAATA TCGTAAGCTT TTTTCTGTGT AGAATCTGCT   1140

GGTTTAACAC CGTCATCTGG AATACCAAGT CGCATATGAC TAACTGGTAT TGCAGCTGCT   1200

ACTAATATGA TTAAACCTAG TAATACTGCC GCAAGTGCAT TTCCTGTAAT AAAATTTAGAC   1260

CATGGCGTAT CAATATCTTT TTTGAATTTA GACTGTAATT TATTCACTTT AATGCGTTtA   1320

TGGAAAATGC TTATTAATGC AGGTAATAAA GTTAAAGCGC TAAGTACTGC AAAAACAACA   1380

CTAATTGCCG AAGCAAATCC CATTACCGCT AAGAAGTCAA TGCCTACTAA TGATAAACCA   1440

CATACTGCAA TTACAACTGT TACACCAGCA AAAACAACTG CACTACCTGC TGTTCCTATT   1500

GCAAGACCAA TGCCTTTAAT GTAATCTGTT TCAGTTTTCA TAACTTGTCG ATATCTGAAT   1560

AAAATAAATA ATGCATAATC GATACCAACT GCTAGTCCAA TCATTACGGC TAATGTCAGT   1620

GTGACATTTG GTATATCGAA TGCATAAGTT AACAAACTGA TAATACCTAC ACCAGAGGCT   1680

AGACCAATCA ATGCACTTAT AATTGGTAAT CCTGCAGCAA TGACTGAACC GAATGTGATT   1740

AACAGTACAA CAAATGCAAC AATAATACCA ACTAGTTCAG AATTACCGCC TACTTCTGTA   1800

```
AAATGACTTT TAACATTATC TCTAGAGCCA TCTTTTAAAG ATGTTTGACT AACGTCATAT    1920

GTGATATCTG CAAATGCAGT TGTTTTATCT TTACTAATTT GCTTATTTTC ATAAGGATCT    1980

GATATTTTAT CAATGTGCTT GTCATCTTTT TTAATATCAT CTAACGTTTT CTTAATATCT    2040

TTAGTAATGT TCGGTTGCAC AATACCATCA TCTTTAGTCG TCTTAAAGAC AACACGTATT    2100

TGTGCCTTTT CACTATCTTG ATTAAAATGT TTTTCAATCT TTTTATTCGT ATCTAACGAC    2160

TCTAATCCTG TCATTTTAAT ATCATTGTCA AATTTCGGTG CATTTGTAGC AAGTGGTATC    2220

AATATTGCAG CTACAATCAC TATCCATGCA ATGACCGCGG ACCATTTATG TTTTGCGATG    2280

AATGTCCCCA TCTTATATAA AAATTTTGCC AAAGTATATT GCCTCCTTTT AAAATCAACG    2340

TTATAGTTTA AATATACAGT GTAGATTATT GTTCGATTAT AGTATCTATC CCCGACCTCT    2400

TAAAGAATCA ATTGGAAAAT TTTGTATATT AAACTACACA CAAAGGAGAA ATGTAGATGA    2460

AAGAGACTGA TTTACGAGTT ATAAAGACAA AAAAAGCATT GTCGAGTAGC TTGCTACAAT    2520

TGTTAGAACA GCAATTATTC CAAACGATTA CTGTCAATCA AATTTGCGAC AACGCACTCG    2580

TACACCGTAC AACATTTTAT AAACATTTTT ATGATAAATA TGATCTTCTA GAGTACTTGT    2640

TCAATCAATT GACTAAAGAC TACTTTGCTA GAGATATCAG TGACCGTCTT AATCATCCAT    2700

TCCAAACGAT GAGTGATACG ATTAATAATA AAGAGGATTT GAGAGAAATC GCAGAATTCC    2760

AAGAAGAAGA CGCTGAATTT AATAAAGTAT TAAAAAATGT CTGCATTAAA ATTATGCATA    2820

ACGATATCAA AAATAATAGA GACCGTATCG ATATTGACAG CGACATCCCA GATAATCTCA    2880

TATTTTATAT TTATGACTCG TTGATTGAAG GTTTTATACA TTGGATAAAA GATGAAAAAA    2940

TTGATTGGCC TGGCGAAGAT ATTGATAACA TTTTCCATAG ATTAATCAAT ATTAAGATTA    3000

AATAGTAGAT GAGAAACTCA TGAGCGTTAC CAACATTCAT AATAAAAACG ATAGTGkACA    3060

CGTTAATGAA TTCGTGTACT ACTATCGTTT TTTATTTTTA TCGTGCTTAT CGCTATTAAA    3120

ACAACTGATA CACAACACAT AAACTATGAA GAAAAAAATA AATCCGCTAT CTAAATGACT    3180

TTGACTCAGT TGTTTAAATG ACCAAATTGC TAATACAATT CCCATTATTA TTGAAATAAC    3240

GTATCTCACA TTCTTATACC TATAATCCTT TTCTAAAAAT ATGGTTGCTA TTACTTAATT    3300

TTTAAAGTTA TAAATAAAAA GAGCCAACCG CAATGGATGG CCCTTGTTCA TTATGAAGCA    3360

TTAGAACATT TCTGAAACAA CCTTTTGTTC TAAGAAGTGT AATAAGTAGT CTGGACTACC    3420

TGTTTTAGCG TCCGTACCTG ACATTTGAA ACCACCAAAT GGATGGTATC CAACAACTGC    3480

TGAAGTACAG CCTCTGTTAA GGTATAAATT GCCTACATCA AATTCGTTTA CCGCTTTAAT    3540

CCAATGCTCG CGATTATTTG TAATCACTGC ACCAGTTAAA CCGTAATCTG TATCATTTGC    3600
```

```
TTCTTCTTGC ATGATTCTAT CTTTAGATTT AAGTCCTGAA ATGATTGTTG GTTCTACAAA        3720

GTAACCTTTT GAATCATCAG TGCCGCCACC TTGTTCTAAT TTACCTTCTT CTTTACCAAT        3780

CTCAATATAA TTTTTAATCT TATCAAATTG TTTTTTATTA ATAACTGGGC CCATATACGT        3840

ATTGTCTACA GTATTGCCCA ACGTTAATTC TTTTGTTAAT TTGATTGATT TCTCTAATAC        3900

TTCGTCATAA ACGTCTTTAT GCACAATTGC ACGTGAACAT GCTGAACATT TTTGACCAGA        3960

AAAACCAAAT GCTGACGTTA CAATAGCTTC TGCTGCCATA TCTGTATCAA TATTTTCATC        4020

AACTACAATG GCATCTTTAC CACCCATTTC AGCGATAACA CGTTTCAAGA AGTTTTGACC        4080

TTCTTGAACA ACGGCACTAC GTTCATAAAT TCTAGTACCT GTCGCACGTG ATCCTGTAAA        4140

TGTAACGAAA TGCGTATCTT TATGATCAAC TAAGTAATCA CCAATTTCTT TCGGATCACC        4200

AGGAACAAAG TTAACTACGC CTTTTGGTAA TCCTGCTTCT TCTAAAATTT CCATTAATTT        4260

ATAAGCGATA TAAGGTGTAT CCTCAGCAGG TTTCAATAAC ACTGTATTAC CTGCCACAAC        4320

TGGTGCTAAA GTTGTACCAG CCATAATCGC AAACGGGAAG TTCCACGGCG GAATTGTAAC        4380

ACCTGTACCA ATTGATTTAT AGAAATATTT ATTGTGTTCA CCTTCACGAT CAAGTACTGG        4440

CTTACCTTGA GCCAAGTCCA TCATTGAACG TGCATAGTAT TCAATAAAAT CAATACCTTC        4500

AGCTGCATCA CCAACTGCTT CATCCCATGG CTTACCTGCT TCATAAACCA TAATTGCTGC        4560

AATTTCCGCT TTTCGACGAC GAATAATTGC CGAAACACGT AACATAAGCT CTGCACGATC        4620

ATTTGCTGAC CATGTTTTCC AAGATTTATA AGCTTCGTTT GCTGCTTTAA ACGCATCTTC        4680

AACATCTTGT TTTGTTGCCT TTGATGCATT TGCAATCACT TGTGATGTGT CTGCAGGATT        4740

GATTGATTTA ATTTTGTCAT CTTTGAAAAT CTTCTCTCCA TTAATCACTA ATGGTATGTC        4800

TTGACCTAAT TCTTTTTCCA CGTCTTTCAA TGCTTTCTTA AACATATCCA CATTTTCTTG        4860

GACTGAAAAA TCGTAACCAG GTTCATTTTT AAATTCTACT ACCATGTACA CTTACCCCCT        4920

ATAAATTTTG AAAGTGGTTT AACCCTTTGA TTTAATGATA TAACATCATT TAAACTCATT        4980

TTACTATGAT TAAGGTTAGT TTTGCAATCG CTTTCATTTT TATGTTTTAT CACTTATTCT        5040

CAAGTATTTT GAAATTGATT GGTTACTTTT TAAAATTTAT ATGGGTCGCA ACTGCTACTT        5100

TATCGTTTCG TCATTTAATG TTTCGGATGG TAGGTCATTA TCAATTTTAC GAACGACTTT        5160

ACAAGGGTTT CCAACCGCTA AGCTGTGTGG CGGAATATCT TTAGTGACAA CACTACCAGC        5220

ACCAATCACA CTGCCTTCTC CAATCGTCAC CCCTGGTAAC ACGGCTACAT GACCGCCAAA        5280

CCAAGTATTA CTGCCAATAT GAATGGGTCC GGCTTTTTCA AAACCTTCAT TTCTATGATG        5340

GAAATTAAGT GGATGTGTCG CTGTGTAGAA TCCACAATTA GGTCCTATAA AAACATTATC        5400
```

```
TCCTAGTTTA ACGTTCCAAC CATAATCTGT ATCAAAAGGA ATCGAAATAC TTACATTGTC       5520

TGTTGTTGTT TGAAATAATT GATCAATTAA TTCCTTTCTT TTATTTGTAG CACTCGGTCT       5580

TGTATGATTT AATTCAAAGC AAATATCTTT CGCTCGTGCA CGTTCATTGA TTAAGTATTG       5640

ATCAAAGTTT GCATCGTACC ATTTTTCTGC TAACATTTTT TCTTTTTCAG TCATTACACC       5700

TTTCAACTCC TAATAACTTA TTTACTTGTT TAAAAGTTAA TCAAATAAAC CTTCGCCTAT       5760

GCAACTAATA CGCTATAACA TTATGAAATC ATGACCTTAT CACCCTTATC TATACAATTC       5820

TCGCATCAAA TACTGCTAAA GTAGTAGATA AATTCAATAC TACAGACGCA TTCATTTTTT       5880

AATCTATTAA CGTACAATGT GAGTAAGAGA AATATAAAGG AGTATGATAG CGATGAGAAT       5940

ATTAATTACA GGCACAGTTG CTATCTTAAT CATTCTAGGT TTGGTCAAAA CGATACAAGA       6000

TTACGAAATG ACAAACGACA CGAGTCGTcA GTTGTCAGAC AACAAAGATG ATGATAAAGT       6060

CATCCATCTT AATAATTTTA AAAATTTACA TGCGAAAGAA TTTAACCCAT CTGATTTCTT       6120

TTAAGTCACC TAAGAATTGC AAATCCAGAA GTCATTTAAG TTTTACCTTT CATTCATACA       ·6180

TCCTTTAATA TTAATTACGA CTTCTTTTAT ATAGATGCTA AGTAGAGAGA TTGTTGTGCA       6240

ATGTTTGCAC GGCAATCTCT CTTTTTCTTT TTAAAATTGG TAAAAGTAAA ACGCAACGAT       6300

TGACTTATAT ACCTATAGGG GGTACATTAG ACGTGTAACA ATGAATCACA GGGAGGCAAT       6360

AATGTGGCTA ATACGAAAAA AACAACATTA GATATCACTG GTATGACTTG TGCCGCATGT       6420

TCAAATCGTA TCGAAAGAA ACTGAATAAA CTTGATGACG TTAATGCCCA AGTGAATTTA        6480

ACTACAGAGA AAGCAACTGT TGAGTATAAC CCTGATCAAC ATGATGTCCA AGAATTTATT       6540

AATACGATTC AACATTTAGG TTACGGTGTC GCTGTAGAAA CTGTCGAATT AGACATTACA       6600

GGTATGACTT GTGCTGCATG CTCAAGCCGT ATTGAAAAAG TGTTAAATAA AATGGACGGC       6660

GTTCAAAATG CAACGGTCAA TTTAACAACA GAGCAAGCTA AAGTTGACTA TTATCCTGAA       6720

GAAACAGATG CTGATAAACT TGTCACTCGC ATTCAAAAAT TAGGTTATGA CGCGTCTATT       6780

AAAGATAACA ATAAAGATCA AACGTCACGC AAAGCTGAAG CGCTACAACA TAAATTGATT       6840

AAGCTTATCA TATCAGCAGT ATTATCTTTA CCACTATTAA TGTTAATGTT TGTACATCTT       6900

TTCAATATGC ATATACCAGC ACTATTTACG AATCCATGGT TCCAATTTAT TTTAGCTACA       6960

CCTGTACAAT TTATTATTGG ATGGCAATTT TATGTAGGTG CTTATAAAAA CTTAAGAAAT       7020

GGTGGCGCCA ATATGGATGT ACTTGTTGCT GTTGGTACAA GTGCAGCATA TTTTTACAGT       7080

ATTTATGAAA TGGTTCGTTG GCTAAATGGC TCAACAACGC AACCGCATTT ATACTTTGAA       7140

ACAAGCGCCG TACTAATTAC CTTAATCTTA TTCGGTAAGT ATTTAGAAGC TAGAGCGAAG.      7200
```

```
TTAAAAGATG GTAATGAAGT GATGATTCCT CTAAATGAAG TACATGTTGG AGATACACTT        7320

ATCGTTAAAC CAGGTGAAAA GATACCTGTT GATGGCAAAA TTATTAAAGG TATGACTGCC        7380

ATCGACGAAT CTATGTTAAC AGGTGAATCT ATCCCTGTTG AGAAGAATGT TGATGATACT        7440

GTAATTGGTT CAACGATGAA CAAAAACGGT ACTATTACTA TGACAGCAAC AAAAGTTGGC        7500

GGGGACACTG CGTTGGCAAA TATTATTAAA GTTGTCGAAG AAGCTCAAAG TTCTAAAGCG        7560

CCGATTCAAC GATTGGCAGA TATTATTTCT GGTTATTTCG TTCCTATCGT TGTTGGTATC        7620

GCACTATTAA CATTTATCGT GTGGATTACT TTAGTTACAC CAGGTACATT TGAACCTGCA        7680

CTTGTTGCGA GTATTTCCGT TCTCGTCATT GCTTGTCCAT GCGCATTGGG ACTTGCTACA        7740

CCAACTTCTA TTATGGTAGG TACTGGTCGC GCTGCTGaAA ATGGTATTTT ATTTAAAGGT        7800

GGCGAGTTTG TTGAACGCAC ACATCAAATT GATACCATCG TTTTAGATAA GACGGGTACC        7860

ATTACAAATG GTCGTCCAGT CGTGACAGAT TATCATGGTG ACAATCAAAC GCTACAACTA        7920

CTTGCTACTG CTGAAAAAGA TTCTGAACAC CCATTGGCAG AAGCCATTGT CAATTATGCA        7980

AAAGAAAAGC AATTAATATT AACTGAGACA ACAACATTTA AAGCAGTACC TGGCCATGGT        8040

ATTGAAGCAA CGATTGATCA TCACCATATA TTGGTTGGTA ACCGTAAATT AATGGCTGAC        8100

AATGATATTA GCTTGCCTAA GCATATTTCT GATGATTTAA CACATTATGA ACGAGATGGT        8160

AAAACTGCTA TGCTCATTGC TGTTAATTAT TCATTAACTG GTATCATCGC AGTGGCAGAT        8220

ACTGTCAAAG ATCATGCCAA AGATGCTATA AAACAATTGC ATGATATGGG CATTGAAGTT        8280

GCCATGTTAA CTGGCGATAA TAAAAACACT GCTCAAGCCA TTGCAAAACA AGTAGGCATA        8340

GATACTGTTA TTGCAGATAT TTTACCAGAA GAAAAAGCTG CACAAATTGC GAAACTACAG        8400

CAACAAGGTA AGAAGGTTGC GATGGTTGGT GACGGTGTAA ATGATGCACC TGCATTAGTT        8460

AAAGCTGATA TCGGTATCGC CATTGGTACA GGTACAGAAG TTGCCATTGA AGCAGCTGAT        8520

ATTACTATTC TTGGTGGCGA CTTGATGCTT ATTCCTAAAG CCATTTATGC AAGTAAAGCA        8580

ACCATTCGTA ATATTCGTCA AAATCTATTT TGGGCATTCG GCTATAATAT TGCCGGTATC        8640

CCTATAGCTG CATTGGGCTT ACTTGCGCCA TGGGTTGCTG GTGCTGCAAT GGCACTAAGT        8700

TCAGTAAGTG TTGTCACAAA CGCACTTAGA TTGAAAAAGA TGCGATTAGA ACCACGCCGT        8760

AAAGATGCCT AGATTCCTTA ATAATGAAGG ATTCGTTGGT GATTCTGAGA TAGGCTAGTG        8820

ATTGGCTCTA TAATGTCGCG GTTTAyaGTt GGATCTTCGC TCCAACTGCA TATATAGTnA        8880

CACTTTTCGC TTGGCGAATT AGTGTATCTT ACCTAATAGc TCCGCCTATT AGGTTCCATC        8940

ATTATTATAA ATAATAAGTA CACTACGGtT TACAGTTGGA TCTTCGCTCC AACTGCATAA        9000
```

```
GAAATTTTAA ATGTTGAAGG TATGAGCTGT GGTCACTGCA AAAGTGCTGT TGAATCTGCA     9120

TTAAATAATA TTGACGGTGT CACTTCAGCT GACGTTAACC TTGAAAATGG TCAAGTAAGT     9180

GTTCAATATG ATGACAGTAA AGTTGCTGTA TCTCAAATGA AAGACGCAAT TGAAGATCAA     9240

GGTTACGATG TCGTTTAATT AGGCAATATT CAACGTCATC AACACCAAAT TAAAAAATCG     9300

AACTGATGAG AATCCCAACA ATCCAAATTA TCTCATCAGT TCGATTTTTA ATTTACTCGT     9360

AACCTAGTAT CTCCAGTCTG CAATACATCT AATGTTGCAT CTAATGCATC GACAATTAGA     9420

TTTTTAACTG CAGCTTCAGT ATAAACGCA  ATATGTGGTG TTAATATGAC ATCTTCCCTG     9480

TCAATCAACG ATTCTAACAA TGGATCGTTC AGTGTTTTGC CCCTTTGATC ACTTGGGAAA     9540

AGTTTGCGTT CAAATTCATA CGTATCAAGT GCTGCACCTT TAATCACACC ATTGTCTAAT     9600

GCGTCTAATA ACGCCTTAGT ATCTACTAAA GAACCTCTCG CACAATTGAC AAATACTGCG     9660

CCCTTTTTAA AATGTTTAAA TAATTCAGCA TTAAATAGAT AATGATTATA TTTCGTTGCA     9720

GGTACATGTA ATGTCACGAT ATCAGCACCT TCAACCGCTT CCTCAATCGT ATCTTTGTAA     9780

TCGACATACG TTGCAATTTT AGCATTAGGA AACGGtCGTA TGCGACCACA TCACTTTGAT     9840

AACCATTGGC AAATATATCG GCTACTACAC GGCCAATTCG ACCTGTACCA ATAACAGCTA     9900

CTTTTAAATC TTTAATGGAT TTCGATAAAA TAGTAGGTTC CCATCTAAAA TCATGcTCCC     9960

GCACTTTCGT TTGAATTTGA TTAAAATGAC GAACCACATT AATAGCCTGG TTCACAGCAA    10020

ACTCCGCAAT TGAATTCGGA GAGTATGACG GCACATTTGA CACAATAAAG TTATACTTGT    10080

TTGCTAACTC CAAATCATAT GTATCAAATC CAGCACTACG TTGTGCGATT TGTTTAATAC    10140

CTAGTTCATT TAATCGTTTA TAAACATGCT CTGATAATGG TATTTGTTGT GATAGCGATA    10200

AGCCATCATA ACCAGCGACA CCTTCAACAT TGTCATCAGT TAATGCTTCT TTAGTAATAT    10260

CTACCTCAAC ATGATGTTTC TCTGCCCACG CCTTGATATA AGGCATATCT TCATCACGTA    10320

CACTCATGAT TTTAATTTTT GTCATTTTAA CATCACCCTT AACTTTATTA TTCATATAAA    10380

TATGCTAGTT CTGTTAATCT TATTGCAGCT TCGTCTAATT TCTGGTCATC TAACGCCAAT    10440

GAAATTCTCA CATAACGATT ACCATTCTCT CCAAATGGTT TCCCTGGAGC AACAAGTATT    10500

GACTTCTCTT GCACTAAAAA TTGCTCAAAT TGCTCGCTGT CATAACCAGG CGGTGTTTCC    10560

AACCATACAT ATATGCCACC TTTAGCATGA ACAAATGGCA AATCAGCTTT TGCAAGCATG    10620

GCTTCGAATC GGTCACGACG TGTTTTAAAT ACATTGCTTT GTTCTTCTAA AAAATCATCA    10680

TAATGATTCA AAGCATATAT TGCGGCATCT TGTAATGCAC CAAACATCCC AGCATTTGTG    10740

TGCGTTTGGT ACTTTTTCAA AGCTTGAATC ATATCTTTAT TACCAACTGC AAAACCGACT    10800
```

```
CCATTTTCCG AAGCAAGTAT ACTAGGATTT TTAGCGTCGA AACCGAAAGC ACCATAAGCA    10920

AAATCATGCA CGATTTTAGT GTCTGTACCT TTAAATTTAG cTATCGCTTC ATCAAAAACT    10980

TCTTTCGTAG CTGTCGATCC AGTTGGATTA TTTGGATACG TTAAATAAAT GAGTTTTGTT    11040

TTATCTATTA TTTGTGAATC AACTTTGGAC CAATCTGGCA AATAATGTGG CGGTTCTAAA    11100

TTAAGCGGGA CTGGCTTGCC ATCAGCTAAA AGTACACCTG CTAAATAATC CGTGTAGCCT    11160

GGATCAGGTA GTAATACATA GTCTCCTGGA TTGATAACAC ATGTTGGTAC TGCCACTAAT    11220

CCATTTTTTG TACCATATAA AATGCATACT TCATCTTCTT TATCTAACGT CACATTATAT    11280

TGTCTTTGAT AAAAATCTAC AATAGCTTGC TTGAACGCTT CTTTACCATG AAAAGCACCA    11340

TATTTTTGAT TTTCAGGAAT AGTTAGTGCT TTTTGAAAAT GATCAATAAT ACCTTGTGGC    11400

GTGGGCCCAT CAGGGATTCC AACTGCCATA TTAATTAATG GCAATGGTCC ATGTTCGATT    11460

TTACGTCCCA TCGTTTTCCC GAAATAACTA TCAGGGATAT TTGCTAATTT GTTAGAGATC    11520

ATCAAATTCC TCCTCTATCA TTAAACATAG CCTGGGCGAC TATCATAATC CTAACAACTT    11580

GTATCACTCT CATTTAGATG GTTACAATGA CATCGCCATT CACCGTTATG TTCAACAGAA    11640

CTTATGACAC ACGTTGTATT GAATGAATTT ATTTTCATTT TAGGTAGGTA TAATATTATT    11700

GTCAATATTA GGAATTTTCA GATTAATATG CACTCAATCG TTATGATTTA ACTGTCATGC    11760

ATATCCGCAT GCGCAACCAG TTAGATATGC TTATATAAAG TATAACGCCC ATCAAGGTAC    11820

GTATTCAAAC GTGAACCTTA ACAGGCGTCA TTCATTGTTA AATAAAACTT CTTAAGCACA    11880

TACTTATTTC ACTATGCCTT TTACGTTCCC CTTATACTTT TCTCACATCT TTCTCTTAGA    11940

CTACTCCCTT ATACGCCCCG CTCAATATCT TTAATCATTT CATCTACAGT TATTTTCGCA    12000

CTCGTTAAGA CAATAGGAAC GCCTGCACCT GGATGCGTAC TTGCACCTGC AAAATATAAA    12060

TCTTTATAAT CTCGCGATAC ATTTTGTGGA CGATAATAAT TACTTTGCGC TAAAGTTGGC    12120

ATTAAACCGA ATGCCGAACC AAATTTCGCA TGATACGTTT GCTCAAAATC ATTTGGCGTA    12180

AAGATTGTTT CTGAAACAAT ATGCGATTTT ATATCTTCAA ATACTTCAAT CGTTGCTAAT    12240

TTACGATAAA TAATTTCCTT TATTTGTTGC GTCAAAGCTT CATCTGACCA ATCGATTCCG    12300

CTACCTGTTT TAAGTTCCGG CGTCGGCATT AGCACATAAA TACCAGTTTT GCCTTCTGGC    12360

GCAAGTGATT TATCAGCGAC CGCTGGTACA TACACATAAA TAGAAGGATC ATATGATAAA    12420

CGTCCCTCAA ATATTTCTTC AATATTGCCT CTAAAGTCAT CTGAAAAAAT AACATTATGA    12480

AGTCTCACTT GATCTGTCAC ATCAATATCT ATACCGATAT ACATTAAAAA TGCTGAACAA    12540

GAGTAATCTA AGTCTGCAAT TTTATGTGGT GGATACTTTT TAATAGGTGC AAAATCTGGC    12600
```

```
ATGTCACCAT TCACTTTTAT CGCATCGGCC CGTTTGAATT TAGGATCAAT AATAATTTGC      12720

TCAATTTCAG CATTTAGTTC AATATTAACG CCTAAGTCTT TATTTAATTG CGCTAGcCCT      12780

TGAGCCATGC CATACATACC GCCTTTAATA AAATGCACAC CAAACATCAT TTCAATCATA      12840

GGAATAATTG AATATAGTGA CGGGCCTCGT TTTGGATCAA TTCCTATGTA TAACGTTTGA      12900

AACGCTAAAA GCTTTTGTAT CTTTTCGTTA TCAATATAAT GTTCAATTAG CTGATCTGCA      12960

TGATTTAACG TTTTTAACTT AGCACCTTGC ACAAGTGACG TCATATTATA AAAGTCACTC      13020

GGTTTGCGAT ACGTTCTTTC TAAGAAATAG CGACGTGCAA TTTCATATTT TTTATAAACA      13080

TCCGTTAAAA AGGACATAAA ACCATGCGTT GAACCAGGTT CTATACTTTC TAGCATTTGC      13140

TGTAATTCAG CTAAATCTGT AGGCACCGTT ATACGATCAT CGTGGTCAAA ATACACATCG      13200

TAAATATAAC GTAATTGTCT CAATTCAATA TAATCTTCAT AATTTTTACC ACACGCTGTA      13260

AAAACATCTT TATAAACATC TGGCATCATG ACAATTGTGG GACCCATATC AAATGTAAAG      13320

CCGTCTTTCT TTAATTGATT CATACGCCCG CCTACATTAT TATTTTTTTC AAATATCGTC      13380

ACTTCATGAC CTTGAGAAGC AATACGGGCT GCCGCTGCTA ATCCTGTGAC ACCTGCACCA      13440

ATTACTGCAA TCTTCATTAT TCAACCACCT ATATTCTATG ATATTTACTA TTTATTTCAT      13500

GAAACAACTT TGCCTTTTTC CTCTTATCCA CAAAAACACG TTCATGTAAT GTATAGTTAG      13560

CCTGTCTCAC TTCGTCCAGT ATTTCAATAT ATATACGTGC TGCTAATTCT ATGATTGGTT      13620

GTGCTTCAAT ACTAAATACT TTGATTTGAT CCATAACATC TTGAAAATCT TTTTCTGCGA      13680

TAGCTGCATA ATATTCCCAT AAGTCAATAT AATGATTATT AACACCATTT TGGTACACTT      13740

CAGCAATATC AACTTCATAT TGCTTTAATC GTTGCTTACT AAAATATATC CGTTCATTGT      13800

CAAAATCTTC ACCGACATCT CTTAATATAT TAAnGGGATC CTCTAGAGTC GACCTG         13856
```

(2) INFORMATION FOR SEQ ID NO: 32:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10088 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

```
ATATATAAAT ATAGATTAAG TATATAGATT AATCAACTTT TTTGGAAGAG CAAATCACGC         60

AATCAACAAA TAATATAAGA AGTTTTTGCG ATAGTTTTAA AATAGCTGTA ATAGAATACT        120

AAATGTGACA AACTTAGAAC TAATATCAAG TGTTGATGTT TTGAATATAA AAATGCTAAT        180
```

```
ATAATTGGTT AATATATGAG TAATTAGAAA ATAGACAAAG GATGACGATT TATGTATATC      300

AATATGAAAG ATTATGGGTT AACAGGCATA AACAAAACTA AAGATACTCG AGCAATACAA      360

CGTGCGTTAA ATCGTGGAAG ATGTAAACCA ACGACAGTTT ATATACCGAA AGGGACGTAT      420

GATATTTGCA AACCATTAAC GATATATGGC AATACAACAC TTTTGTTAGA TAATGAAACT      480

ATTTTACGCC GATGTCATTC TGGTCCTTTA TTAAAAAATG GTCGTCGCTT TGGTTTTTaT      540

CGTGGTTATA ATGGACACAG TCATATTCAT ATTAAAGGCG GCAAGTTTGA TATGAATGGT      600

GTATCGTATC CTTATAACAA TACAGCTATG TGCATTGGGC ATGCTGAAGA TATTCAATTA      660

ATAGGTGTGA CCATTAAGAA TGTAGTGAGT GGTCATGCAA TTGATGCTTG TGGGATTAAC      720

GGACTCTATA TTAAAAGCTG TTCATTTGAA GGATTCATAG ACTATAGTGG CGAACcTTTT      780

ATTCTGAAGC AATACAATTA GACATTCAAG TACCTGGTGC TTTTCCAAAA TTCGGAACgA      840

CAGATGGTAC GATAACGAAA AATGTCATTA TCGAAGATTG TTATTTTGGA CCTTCAGAAT      900

TGCCCGAAAT GGGAAGTTGG AATCGTGCTA TTGGCTCACA TGCAAGTAGA CATAATCGAT      960

ACTATGAGAA TATTCATATT AGAAATAATA TATTTGAAGA TATACAAGGT TATGCATTAA     1020

CTCCCTTGaA GTATAAAGAT GCTTTCATTA TTAATAATAA GTTTATTAAC TGTGaGGGTG     1080

GCATTAGATA TTTAGGAGTT AGAGATGGTA AAAATGCAGC AGATGTGaTG ACAGGaAAAG     1140

ACTTAGGTTC CCAAGCAGGC ATAAATATGA ATATAATTGG AAATGAATTT AAAGGATCAA     1200

TGTCTAAAGA TGCGATACAT GTACGTAATT ATAATAATGT TAAACATAAA GATGTATTAA     1260

TCGTTGGGAA TACATTCAAT AATTCGACTC AATCAATTCA TTTAGAAGAT ATTGATACAG     1320

TGTTTTTAAG TCCTGTTGAA GCGGGTATTC AAGTTACTAC AATCAATGTA GATGAAATAA     1380

AAAAGTAAAA AGTTTCGCAT GACATTAGGA TTAAGAATAG TAGATAATTT TTGAAAGCGC     1440

ATTeATAAAA CGGTATAAAT ATGCTATAAT AAACCCAATT ATCTGATAAA AGGGGTATTT     1500

TGACGGTAAT GATAATACAA GATAGACAAC TTTCTATACT CTAATATAGT GAGTTGAAGT     1560

AGCTTGTCAT AATCATCATG AGGGGGAAAT TTATGGCTTA TTTCAATCAA CATCAATCAA     1620

TGATATCGAA AAGGTATTTA ACATTCTTTT CAAAATCAAA GAAAAGAAA CCGTTTAGTG     1680

CGGGACAACT TATTGGACTA ATATTAGGTC CATTACTTTT CCTATTAACA TTATTATTCT     1740

TTCATCCACA AGACTTACCT TGGAAAGGCG TCTATGTTTT AGCGATTACT TTATGGATTG     1800

CGACTTGGTG GATTACTGAA GCAATTCCTA TTGCAGCAAC GAGCTTATTA CCAATTGTGT     1860

TATTACCATT AGGTCATATA CTTACACCAG AACAAGTATC ATCCGAATAT GGCAATGATA     1920

TTATCTTTTT GTTTTTAGGT GGATTTATTT TGGCAATTGC AATGGAAAGA TGGAATTTAC     1980
```

```
TTGGATTCAT GGTGGCAACA GGATTCTTAT CTATGTTTGT ATCGAACACT GCAGCTGTAA        2100

TGATTATGAT TCCGATTGGT TTAGCAATTA TTAAGGAAGC ACATGATTTA CAAGAAGCCA        2160

ATACGAATCA AACAAGTATT CAAAAGTTTG AAAAATCTCT AGTTTTAGCA ATTGGCTATG        2220

CAGGTACGAT TGGTGGCTTG GGTACATTAA TCGGAACCCC GCCATTAATT ATTTTAAAAG        2280

GACAATACAT GCAACATTTT GGACATGAAA TTAGTTTTGC TAAATGGATG ATTGTAGGGA        2340

TTCCAACGGT CATTGTTTTG TTAGGTATTA CTTGGCTCTA TTTAAGATAT GTTGCGTTTA        2400

GACATGATTT GAAATATTTa CCTGGTGGTC AGACGTTAAT TAAACAAAAG TTAGACGAGC        2460

TTGGCAAAAT GAAGTATGAA GAAAAGGTAG TACAAACTAT CTTTGTACTT GCTAGCTTAT        2520

TATGGATTAC AAGAGAGTTT CTTCTGAAAA AATGGGAAGT TACGTCATCT GTTGCAGATG        2580

GTACGATTGC TATTTTTATA TCAATATTAT TATTTATTAT TCCAGCTAAA AATACTGAAA        2640

AACATCGCCG TATCATTGAC TGGGAAGTTG CAAAAGAGCT CCCTTGGGGT GTATTAATTT        2700

TATTTGGTGG CGGTTTAGCA TTAGCGAAAG GTATTTCTGA AAGTGGTTTA GCAAAATGGT        2760

TAGGCGAACA GTTGAAATCA TTAAATGGTG TTAGTCCGAT TCTTATTGTA ATTGTCATAA        2820

CAATCTTTGT CTTATTTTTA ACTGAAGTGA CATCTAATAC TGCAACTGCA ACGATGATTT        2880

TACCGATTTT AGCAACGTTG TCTGTTGCTG TTGGAGTGCA TCCATTACTA CTTATGGCAC        2940

CTGCAGCTAT GGCGGCTAAC TGTGCATACA TGTTACCAGT AGGGACACCA CCGAATGCAA        3000

TTATCTTTGG TTCTGGTAAA ATATCTATCA AACAAATGGC ATCAGTAGGA TTCTGGGTAA        3060

ACTTAATCAG TGCAATAATT ATTATTTTAG TCGTGTATTA TGTAATGCCT ATAGTTTTAG        3120

GTATTGATAT AAATCAACCA CTGCCATTGA AATAGTAATT GCAGATTAGA ACGAAAAATA        3180

AAAGGTTACA TTAGCAATTG CTTGGACGAG TGGTAACGAA ACGTATACCG CAGCATCGTG        3240

TAAΣAACAAT ACAAACAAAA GAAAGTCAAC CAAGGATGGA TTCCTATTTT AATCCTTGGT        3300

TGACTCTTTA TTTTATTTAA ATTGTAGAAC CTAGAAAATA AAGTTTAATT AAAAGCACCA        3360

ATCATTTCTA CTTTGAAATC TAAGGTTTCT AAAATAGCAA TGACTTTCTT TATATCGGTT        3420

GTAATTGCAG AATCAGCCTG AACGAAAAAT CGATACATAC CTAATTGTGT TTTTAAAGGA        3480

CGAGACTCAA TCCAGGATAA ATTAATATTA AACAAAGCAA ATGTATTAAG CACACTTGCT        3540

AACAACCCAG GTTTATCATG CATTGGTGTA ATTAAAAACA TCAATGATGT CGCATTTTGA        3600

TCAAATTGCT GCTGATTTTT TATAACTAAA AAACGTGTCA CGTTATGTGG ATAGTCTTCA        3660

ATATGTGTAT CAATAGGTGT AAAACCATAA GctTCGCCAC TACCTAAAGG TGCAATTGCT        3720

GCAACGCCAT TTTCAATTTT AGTCAAACTT TGAATTGTAC TGTCGACATA ATCATAGTCA        3780
```

317

```
TTTTTAATAT CAGAAATGGA ATCTGTTCCA TTACCATATA ATGCAAAGTT AATATCTAAA       3900

CGTATTTCAC CGTGTGCAAA GACATCTTGC TGTGCAAGTG CATCTGCCAC AATGTTGATT       3960

GTTCCTTCTA TAGAATTTTC AATAGGGACA ACACCAATCG ATGTGTCATC ATCTGCAACT       4020

GCCTTGATGA CTTCAAATAA ATTTGACTTT GGTTGAAAAG TTGCTTCATT TTCAGAAAAA       4080

TACTGACGAC AAGCCAAATA TGAAAATGTA CCTTTAGGGC CTAAATAATA TAATTGCATA       4140

TGCTACACCT CTACTAACTT AATGATGGAA AGGGCACTGG TTAGCATTTG ATTCTTTCTT       4200

TTTATAGAAA AAGTTTGGAT CTTTTACTGT ATTGTCATAT CCGTGATGAT AATTTGACGT       4260

CAATGTTGGA GATAATGGCG GTGCTAGCCA AGACCATTTT CCGGTAACTT GACGACCTTG       4320

TTGTGCTTCG TTACGTTCGA ATAGTTCGAA TTGCTTTGCA GCGGTCAAAT GATCGACAAT       4380

TGATACGCCT TCTTTTTTAA AGGAATGATA CACAGCATAG TTCAATTCAA CAAGTGCTCG       4440

ATCTTTATTA AATGAATTAT TTTTAAGTGT ATCAAATTCA AACGCATCTG CAACTTTTTC       4500

TAGTAAATTG TAACGGTAAT CATCAATAAA GTTACGTACG CCAATTTCAG TTACCATATA       4560

CCAACCGTTA AAGGGTGCAG TTGGATATAC AATGCCACCG ATTTTTAAGT CCATATTGGA       4620

AATGATAGGG ACTGCATACC ATTTTAAGTT CAATTTTCTT AATTTTGGAT AATGATTATG       4680

TTCAATAGGT ACTTCTTTAA TTAATGAAGT AGGATATTCG TAAAATTTAA CTGACTCATT       4740

AGGTAATTGG TAAATCAGTG GTAACACGTC AAAATTAGTA CCTTTTCCTT TCCAACCTAA       4800

GTGATTTGCT AAGCGTGTAA CTTCTTTTTC AGCAGGATCA CCACAATTGT CATAGCCAGC       4860

ATAGCGAATT AATTGATTGT TGAAAATTTT AGGTCCATCC TTTGGAGCAT ATATAGTAAT       4920

ATACGGCTTT AATTTACCTT CATTTGTAGC CTGTGTAATA TGATAAGTAA TTGATGATAA       4980

GAACGATGCT TCGTCAGTAA CATCTCTTGC ATCAATGACA TTTAACGAAT CCCAAAATAA       5040

ACGACCAATG CAACGATTTG AATTACGCCA AGCCATTTTA GCACCATAAA TAAGTTCTTC       5100

TTCTGTATGT GTATATGTCC CAGTTTCTTT TATTTCTAGT TCAATGTCAT GTAAACGTTT       5160

ATTGATAATT TGCGTTTCAT AATGACACTC TTTATACATG TTTTCTATGA AAGCTTGAGC       5220

CTCTTTAAAT AACATTAACA ACACCTCGCT TTATATTATA GTCTACATTA TTAAAATACT       5280

CTTAAAAATT ATGTATATGT CATTAAATTG TTGGTTGATT TTAATTAAAA GTATGGAAAT       5340

TAAGGGGCTC TTATGTATAT AAAAAAATGA ATTATGATAA AATGTAAGAA AATATTTAGG       5400

TCGATTGGAG AGATACAAGT GTACCAATTA GAAGACGACA GTTTAATGTT ACATAATGAC       5460

TTATATCAAA TAAATATGGC TGAAAGTTAT TGGAATGATA ATATTCATGA AAAAATGGCT       5520

GTATTTGATT TGTATTTTAG AAAAATGCCA TTTAATAGTG CTATGCTGT TTTTAATGGT        5580
```

```
TTAAAGTCTA TTGGCTACAA GGATGATTTC TTATCATATT TAAAAGATTT AAAAATTCACA    5700

GGCAGCATCC GTTCGATGCA AGAAGGCGAA TTATGCTTTG GTAACGAACC ATTGTTACGC    5760

GTAGAAGCAC CATTGATTCA AGCGCAATTA ATAGAAACAA TTTTATTAAA CATTGTAAAT    5820

TTCCATACAT TAATTACAAC AAAGGCTAGC AGAATTCGTC AAATTGCATC AAATGATAAA    5880

TTAATGGAGT TTGGTACACG TCGTGCGCAA GAAATTGATG CAGCATTGTG GGGCGCTAGA    5940

GCTGCTTACA TCGGGGGCTT TGATTCTACA AGTAATGTTA GGGCGGGGAA ATTATTTGGT    6000

ATACCTGTGT CTGGTACACA TGCACATGCA TTTGTCCAAA CTTATGGAGA CGAATATGTT    6060

GCCTTCAAAA AATATGCTGA AAGACATAAA AATTGTGTGT TCCTAGTAGA TACATTCCAT    6120

ACTTTAAAAT CTGGCGTGCC AAATGCAATA AAAGTTGCAA AAGAATTAGG TGACAAAATT    6180

AACTTTGTAG GTATTCGATT AGATTCTGGA GATATCGCTT ATTTATCTAA AGAGGCAAGA    6240

CGTATGCTTG ATGAAGCAGG ATTTACTGAA ACTAAAATTA TCGCGTCTAA TGATTTGGAT    6300

GAAGAAACGA TTACGAGTTT GAAAGCACAA GGTGCAAAAG TAGATTCTTG GGGCGTTGGT    6360

ACAAAGCTGA TTACAGGATA CGATCAACCA GCATTAGGTG CAGTATATAA ACTTGTAGCT    6420

ATTGAAAATG AAGATGGTTC ATATAGTGAT CGTATTAAAT TATCAAATAA CGCTGAAAAG    6480

GTTACGACGC CAGGTAAGAA AAATGTATAT CGCATTATAA ACAAGAAAAC AGGTAAGGCA    6540

GAAGGCGATT ATATTACTTT GGAAAATGAA AATCCATACG ATGAACAACC TTTAAAATTA    6600

TTCCATCCAG TGCATACTTA TAAAATGAAA TTTATAAAAT CTTTCGAAGC CATTGATTTG    6660

CATCATAATA TTTATGAAAA TGGTAAATTA GTATATCAAA TGCCAACAGA AGATGAATCA    6720

CGTGAATATT TAGCACTAGG ATTACAATCT ATTTGGGATG AAAATAAGCG TTTCCTGAAT    6780

CCACAAGAAT ATCCAGTCGA TTTAAGCAAG GCATGTTGGG ATAATAAACA TAAACGTATT    6840

TTTGAAGTTG CGGAACACGT TAAGGAGATG GAAGAAGATA ATGAGTAAAT TACAAGACGT    6900

TATTGTACAA GAAATGAAAG TGAAAAAGCG TATCGATAGT GCTGAAGAAA TTATGGAATT    6960

AAAGCAATTT ATAAAAAATT ATGTACAATC ACATTCATTT ATAAAATCTT TAGTGTTAGG    7020

TATTTCAGGA GGACAGGATT CTACATTAGT TGGAAAACTA GTACAAATGT CTGTTAACGA    7080

ATTACGTGAA GAAGGCATTG ATTGTACGTT TATTGCAGTT AAATTACCTT ATGGAGTTCA    7140

AAAAGATGCT GATGAAGTTG AGCAAGCTTT GCGATTCATT GAACCAGATG AAATAGTAAC    7200

AGTCAATATT AAGCCTGCAG TTGATCAAAG TGTGCAATCA TTAAAAGAAG CCGGTATTGT    7260

TCTTACAGAT TTCCAAAAAG GAAATGAAAA AGCGCGTGAA CGTATGAAAG TACAATTTTC    7320

AATTGCTTCA AACCGACAAG GTATTGTAGT AGGAACAGAT CATTCAGCTG AAAAATATAAC   7380
```

```
TAAACGACAA GGTCGTCAAT TATTAGCGTA TCTTGGTGCG CCAAAGGAAT TATATGAAAA    7500

AACGCCAACT GCTGATTTAG AAGATGATAA ACCACAGCTT CCAGATGAAG ATGCATTAGG    7560

TGTAACTTAT GAGGCGATTG ATAATTATTT AGAAGGTAAG CCAGTTACGC CAGAAGAACA    7620

AAAAGTAATT GAAAATCATT ATATACGAAA TGCACACAAA CGTGAACTTG CATATACAAG    7680

ATACACGTGG CCAAAATCCT AATTTAATTT TTTCTTCTAA CGTGTGACTT AAATTAAATA    7740

TGAGTTAGAA TTAATAACAT TAAACCACAT TCAGCTAGAC TACTTCAGTG TATAAATTGA    7800

AAGTGTATGA ACTAAAGTAA GTATGTTCAT TTGAGAATAA ATTTTTATTT ATGACAAATT    7860

CGCTATTTAT TTATGAGAGT TTTCGTACTA TATTATATTA ATATGCATTC ATTAAGGTTA    7920

GGTTGAAGCA GTTTGGTATT TAAAGTGTAA TTGAAAGAGA GTGGGGCGCC TTATGTCATT    7980

CGTAACAGAA AATCCATGGT TAATGGTACT AACTATATTT ATCATTAACG TTTGTTATGT    8040

AACGTTTTTA ACGATGCGAA CAATTTTAAC GTTGAAAGGT TATCGTTATA TTGCTGCATC    8100

AGTTAGTTTT TTAGAAGTAT TAGTTTATAT CGTTGGTTTA GGTTTGGTTA TGTCTAATTT    8160

AGACCATATT CAAAATATTA TTGCCTACGC ATTTGGTTTT TCAATAGGTA TCATTGTTGG    8220

TATGAAAATA GAAGAAAAAC TGGCATTAGG TTATACAGTT GTAAATGTAA CTTCAGCAGA    8280

ATATGAGTTA GATTTACCGA ATGAACTTCG AAATTTAGGA TATGGCGTTA CGCACTATGC    8340

TGCGTTTGGT AGAGATGGTA GTCGTATGGT GATGCAAATT TTAACACCAA GAAAATATGA    8400

ACGTAAATTG ATGGATACGA TAAAAAATTT AGATCCGAAA GCATTTATCA TTGCGTATGA    8460

ACCTCGAAAC ATACATGGTG GATTCTGGAC TAAAGGCATT CGTCGTAGAA AGCTTAAAGA    8520

TTATGAACCA GAAGAACTGG AAaGTGTAGT AGAaCATGAA aTTCmAAGTA AaTGAGAaTG    8580

AAmCAATtGC TGATTGTTTG TCACGAATGA AAtGCAAGGG TATATGCCGG TAAAACGTAT    8640

TGAAAAACCC GTGTTTCAAG AGCAAAAAGA TGGCACGGTT GAAGTATCAC ATCAAGAAAT    8700

CGTTTTTGTA GGTAAGAAAA TCCAATAACA TAATCCAATT TAAATAAAGA CTATTTGAAG    8760

AGGAAAGGCT ATTCAAAGTT TGAGTAATTT TACTTTGAAT AGCCTATTTG TTTATACATG    8820

CAAGATGCTC GATCCATATT GTATGAGAAA CCCCCAGCAA GCTATATAAA GCATATGCTG    8880

GGGGTTCTTA ATATTTTAAA AATTATTGTT AGATTATATA TATCGTCGCT TTTTCTAAAA    8940

CAATCTCATC GCATGAAATT TTTTCTTCCT AGAGACCTTT AATAAGATTA ATAGTTTACT    9000

TAATCATATC TAGATAGTCT TATGACTTAT GCTTAATGAA AGTCATTCTA GGAGAAGTTC    9060

CCAAAGCTTC TGTGTTCATA ATTGTTAGTA GTATTTTATT ATCATTTGGT ATAAATATTT    9120

CAATAACAAT TGAGCTATTA TTTTTATTAT ATAATGTGAG TTGTTTGTGT TCTGTATTTA    9180
```

CATTTAAATC TTGAGGATGC CATTCTCCCT CAATAATATT AAGATAATAC TTAGCCTCTG   9300

AATTACATTT GAATTTATCA ATACTAAATA ATTCAATTTG TTCCATAATA TTATTTACCT   9360

TTCTAAAATA CAAATTTTAA TAACCATAAA TAGATGAATA CCATCGATAA TGGTCGCCAT   9420

TGGATACTGG AATAACATTG TTTTTAGCAT CTTGAGTCAT AAAACCATTA TCCCATGGAT   9480

TCCATATAAT TATAACCTCT TGTCCATTAT CTAATTTAGC GTTCCCAACA ACTGCCATGG   9540

CATGCCCTGC GTGCATACCA TTTCTTGATT CTACTCTACT ACCTAAAACA GCAATTCCTT   9600

TATTATTTTT AGTAAGATTG TCAACTTCAT TATATGTAGT CATTCTATTA AGAAGTTGTG   9660

GACTTCTTCC CTGAGTTTGT CCAAAATAAA TCATCTCTCT TGGCGTTAAA CCAGTAAATT   9720

GGAATCGTTG TCCTTGTAAG TTTGGGTGTA AAAATCTCAT CACAGCTTCT GCATGATATT   9780

TGTTAGTATT ATAAGTCGCA TTTAGTAATT CAGACATCGT ATAGCCTGCA CACCAACCAT   9840

TGTTACCTTG AGTTTCTCTT ATCTTGAAAT TCTCAAGTTT ATTTATATAT TGsTCGTTGT   9900

AAGTATAATT ATTACTTTTA AATTGACTAG TTGGCATAGT GACAGAAGCT TTTTGCTTTA   9960

GTTGCGTTAC ATTATTGCCA GTAGGTATAC TCTCAGTCTT TnTnAACTnT nTATCTTCTA   10020

GACGTGGTGT TTTTAGTACT AGTTTAGCTT TATGATTTTG AGTACCACAT AGTAACCTTT   10080

TGAGTTGT   10088

(2) INFORMATION FOR SEQ ID NO: 33:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7563 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGGAAACGnA CCCnATGCGT ATGCTTGACG TGCCAAAATT AAATACGAAG TTCATAGCTT   60

TGAGGTACCA GAAGAACATT TATCTGGTCA AGAAGTCGCA GnACTCATAC AAGCAAATGT   120

TAAAACAGTA TTTAAAACGC TTGTTCTAGA AAATACAAAA CATGAACATT TTGTATTTGT   180

TATCCCAGTA AGTGAAACTT TAGATATGAA AAAGGCAGCT GCTTTGGTTG GAGAGAAGAA   240

ATTGCAGCTT ATGCCTTTAG ATAATTTGAA AAATGTAACG GGATACATTC GTGGTGGGTG   300

TTCGCCTGTT GGTATGAAAA CATTGTTTCC AACAGTCGTT GACAAATCGT GTGAAAATTA   360

TAGTCATATC AGTGTGAGTG GTGGGCTTCG AACAATGCAA ATCACAATAG CTGTTGAGGA   420

TTTGATTACA ATAACTAAAG GCAAAATTGG AGCAGTTATC CATGAATGAT TAATAACAAC   480

```
TGCCACACTC CTTTTTGATT GAATTAGCAT TTTACGATCA TAAACAGTCA TTATAATTGA      600

GTATTTGAAC ATAAAAATGT AATTTTATCG TAACAATTTG AGTGTTTGTG ATTGTTTTTG      660

GTAATTTATG ATTGAAAAGT GAAAGCGTAC TCATTATAAT ACAAAGTGAG ATGGGGTGAT      720

GATGATAATT ACTGaAAAAA GACACGAGTT AATATTAGAA GAACTTTCGC ACAAAGATTT      780

TTTGACTTTA CAAGAATTAA TAGATCGAAC TGGTTGCAGT GCTTCAACAA TACGArGAGA      840

TTTATCTAAA CTACAACAAT TAGGGAAATT GCAACGTGTG CATGGTGGTG CAATGTTAAA      900

AGAAAATCGT ATGGTTGAGG CGAATTTAAC TGAAAAATTA GCAACGAATC TTGATGAAAA      960

GAAAATGATT GCTAAAATAG CAGCTAATCA AATCAACGAT AATGAATGCT TATTTATCGA     1020

TGCTGGTTCA TCTACATTGG AGCTAATTAA ATATATTCAA GCGAAAGATA TCATTGTGGT     1080

AACCAATGGT TTAACACATG TAGAAGCTTT ACTTAAAAAA GGTATTAAAA CAATTATGCT     1140

AGGTGGTCAA GTTAAAGAAA ATACACTTGC TACGATTGGT TCTAGTGCTA TGGAGATATT     1200

AAGACGATAT TGTTTCGATA AAGCTTTTAT CGGGATGAAT GGATTAGATA TTGAACTTGG     1260

ATTAACTACT CCCGATGAGC AAGAGGCATT AGTTAAACAA ACAGCAATGT CATTAGCCAA     1320

TCAATCATTT GTACTTATAG ATCATTCTAA GTTTAATAAA GTATATTTTG CTCGTGTACC     1380

TTTGCTAGAA AGTACGACAA TCATCACATC TGAAAAAGCA TTAAATCAAG AATCGTTAAA     1440

AGAATACCAA CAAAAGTATC ACTTTATAGG AGGGACTTTA TGATTTATAC AGTGACTTTC     1500

AATCCTTCAA TTGACTATGT CATTTTTACG AATGATTTTA AAATTGATGG TTTGAACAGA     1560

GCAACAGCAA CATATAAATT CGCTGGGGGG AAAGGTATTA ATGTCTCGCG CGTCTTAAAG     1620

ACATTGGATG TTGAGTCAAC TGCCTTGGGA TTTGCAGGTG GATTTCCTGG GAAATTCATT     1680

ATAGATACAT TAAATAACAG TGCAATTCAA TCGAATTTTA TTGAAGTTGA TGAAGATACA     1740

CGTAJTAATG TGAAATTAAA AACAGGACAA GAAACAGAAA TCAATGCACC GGGTCCTCAT     1800

ATAACGTCAA CACAATTTGA ACAACTGTTA CAACAAATTA AAAATACAAC AAGCGAAGAT     1860

ATAGTTATTG TTGCTGGAAG TGTACCAAGT AGTATTCCAA GCGATGCGTA TGCGCAAATT     1920

GCACAAATTA CAGCACAGAC AGGTGCTAAA TTAGTAGTCG ACGCTGAAAA AGAATTGGCT     1980

GAAAgCGTTT TACCATATCA TCCACTATTT ATTAAACCTA ATAAAGATGA ATTAGAAGTG     2040

ATGTTTAATA CAACAGTGAA CTCAGACACA GATGTTATTA AATATGGTCG TTTGTTAGTT     2100

GATAAAGGTG CGCAATCTGT TATTGTCTCG CTTGGCGGTG ATGGTGCTAT TTATATTGAT     2160

AAAGAAATCA GTATTAAAGC AGTTAATCCA CAAGGGAAAG TGGTTAATAC AGTTGGCTCT     2220

GGTGATAGTA CAGTTGCAGG CATGGTGGCT GGAATTGCTT CAGGTTTAAC GATTGAAAAA     2280
```

```
CGGGACGCTA TAGAAAAAAT AAAATCACAA GTTACGATTA GCGTACTTGA TGGGGAGTGA      2400

AAATAATGAG AGTAACAGAG TTATTAACAA AAGATACAAT AGCAATGGAT TTAATGGCAA      2460

ATGACAAAAA TGGTGTTATT GATGAGTTAG TAAATCAATT AGACAAAGCA GGTAAATTAA      2520

GTGATGTCGC GTCATTTAAG GAAGCGATTC ACAATCGAGA ATCACAAAGT ACAACTGGTA      2580

TCGGCGAAGG TATTGCCATT CCACATGCCA AAGTGGCCGC AGTTAAGTCA CCAGCTATTG      2640

CGTTTGGTAA ATCTAAAGCA GGCGTAGATT ATCAAAGTTT GGATATGCAA CCAGCACACT      2700

TATTCTTTAT GATTGcAGcG CCAGAAGGTG GCGCCCAAAC ACATCTAGAT GCTTTAGCTA      2760

AGTTGTCTGG TATTTTAATG GATGAAAATG TACGTGAGAA ATTATTACAT GCTTCATCAC      2820

CTGAAGAAGT ACTAGCGATC ATAGATGAGG CTGATGATGA AGTGACAAAA GAAGAAGAGG      2880

CAGAAGCTGA AGCACAACAA GTTGCAACTG CAGAACAATC ATCTAAACAA TCTAATGAGC      2940

CATATGTGTT AGCAGTAACT GCTTGTCCAA CAGGTATTGC ACACACATAT ATGGCACGTG      3000

ATGCATTGAA AAAGCAAGCG GATAAAATGG GTATTAAAAT TAAAGTAGAA ACGAATGGTT      3060

CAAGCGGCAT TAAAAACCAT TTAACTGAAC AAGATATTGA AAATGCAACA GGTATCATTG      3120

TTGCTGCTGA TGTTCATGTT GAGACGGATC GCTTCGATGG TAAAAATGTC GTAGAAGTAC      3180

CAGTAGCAGA TGGTATTAAA CGCCCAGAAG AATTAATTAA TAAAGCATTA GATACAAGTC      3240

GTAAACCTTT TGTTGCCCGT GATGGTCAAA GAAAAGGTAA CTCAAATGAC AGTCAAGAAA      3300

AATTAAGCCC AGGTAAAGCA TTCTATAAAC ACTTAATGAA CGGTGTTTCT AACATGTTGC      3360

CACTTGTAAT ATCTGGTGGT ATTTTAATGG CAATTGTATT TTTATTTGGA GCAAATTCAT      3420

TTAATCCAAA AAGCTCAGAG TACAATGCGT TTGCAGAGCA GCTTTGGAAC ATTGGTAGTA      3480

AAAGTGCATT CGCGTTAATC ATTCCAATTT TATCTGGATT CATTGCACGT AGTATTGCGG      3540

ATAAACCTGG TTTCGCTTCA GGTCTTGTAG GTGGTATGTT AGCAATTTCA GGTGGTTCAG      3600

GATTTATTGG TGGTATTATT GCAGGTTTCT TAGCAGGTTA CTTAACACAA GGTGTTAAAG      3660

CCATGACACG TAAGTTACCA CAAGCATTAG AGGGATTAAA GCCAACATTA ATTTATCCAC      3720

TATTAACAGT GACGGCTACA GGCTTATTGA TGATTTATGC CTTTAATCCA CCAGCATCTT      3780

GGTTAAATCA TTTGTTATTA GATGGATTAA ACAATTTATC AGGTTCTAAT ATTGTATTAT      3840

TAGGTTTAGT TATTGGCGCT ATGATGGCGA TTGATATGGG CGGTCCATTC AACAAAGCGG      3900

CATATGTTTT TGCAACAGGT GCGTTGATTG AAGGTAATGC AGCACCAATT ACAGCTGCAA      3960

TGATTGGTGG TATGATTCCA CCGTTAGCAA TTGCGACAGC GATGTTAATT TTTAGACGTA      4020

AATTTACAAA AGAACAACGT GGTTCAATTA TCCCTAACTA TGTGATGGGT ATGTCATTTA      4080
```

```
TGATTGGTTC AGGTATAGGT GGCGCAATTG CTTTAGGCTT AGGTTCACGA ATTACTGCGC    4200

CACATGGTGG TATTATTGTA ATTGTTGGTA CTGATGGTGC ACACTTACTT CAAACTCTTA    4260

TTGCACTTCT AGTTGGCACA TTAGTTTCAG CATTAATTTA CGGTTTAATC AAACCAAAGT    4320

TAACTGAAAC AGAAATCGAA GCTTCAAAAT CAATGGACGA GTAGTTTTAA TGATGTAAAA    4380

TGATTGTTAG CAAAGAGCTT CATATTAAGT TGTATGTTCA ATGAATATAT GTTAGTTTTA    4440

TATATCGTGT TAACGGTAGC TTATACAAAG CTGTAAAAAC ACTTTCTATT AATTCAGTTT    4500

TTATGAATTG ATATGAAAGT GTTTTTATTT TTAGATAAAT GAATGAAGAA ATAGACACCA    4560

CAAATGTATA GACTTTTTTA ATATTTTGCA AAAAGTTATG CCAAACGAAG CAGATATAGT    4620

AAAATATGAG TGTCTTAAAG TGAAAATTTA TAAATAAAGA AGGGTTTATA CGTGTCAGAA    4680

TTAATTATAT ATAACGGCAA AGTTTATACT GAAGATGGCA AAATCGATAA TGGTTACATT    4740

CATGTGAAAG ATGGACAGAT TGTTGCAATT GGAGAAGTGG ATGATAAAGC AGCAATTGAT    4800

AATGATACGA CAAATAAAAT TCAAGTGATT GATGCTAAAG GTCATCATGT ATTACCAGGT    4860

TTTATTGATA TACATATTCA TGGTGGTTAT GGTCAAGATG CAATGGATGG TCATACGAT     4920

GGCTTAAAAT ATCTATCCGA AAATTTGTTG TCTGAAGGGA CGACATCATA CTTGGCCACT    4980

ACAATGACGC AATCGACTGA TAAAATAGAT AATGCACTTA CAAATATTGC TAAATATGAA    5040

GCGGAgCAAG ATGTTCACAA TGCAGCGGAA ATTGTAGGTA TACATTTAGA AGGACCATTT    5100

ATATCTGAAA ATAAAGTTGG TGCTCAACAT CCGCAATACG TTGTACGCCC ATTTATCGAT    5160

AAAATTAAAC ATTTTCAAGA GACTGCTAAC GGATTAATAA AGATTATGAC GTTTGCACCT    5220

GAAATTGAAG GTGCAAAAGA AGCGCTTGAA ACGTATAAAG ATGACATTAT TTTTTCAATT    5280

GGTCATACAG TAGCAACATA CGAAGAAGCA GTTGAAGCTG TTGAGCGAGG AGCTAAACAT    5340

GTCACGCATT TATATAATGC AGCGACGCCA TTCCAACATA GAGAACCAGG TGTTTTTGGA    5400

GCAGCATGGT TGAATGATGC TCTACATACC GAAATGATTG TTGATGGCAC TCATTCTCAT    5460

CCGGCATCGG TTGCAATTGC TTACCGTATG AAAGGTAATG AACGTTTTTA TTTAATTACC    5520

GATGCAATGC GTGCAAAAGG TATGCCTGAA GGAGAATATG ATTTGGGTGG ACAAAAAGTA    5580

ACTGTTCAAT CGCAACAAGC ACGTCTTGCA AATGGTGCGC TTGCTGGTAG TATTTTAAAA    5640

ATGAATCATG GGTTACGTAA CTTAATATCA TTTACAGGTG ATACATTAGA TCATTTATGG    5700

CGAGTAACAA GTTTAAATCA AGCCATTGCA TTAGGTATCG ATGATAGAAA AGGTAGTATT    5760

AAAGTAAATA AGGATGCAGA TCTTGTTATT CTAGATGATG ATATGAATGT AAAATCTACA    5820

ATAAACAAG GCAAGGTTCA CACATTTAGC TAATAAATAA TCATAATTAA ATGTATGCAA     5880
```

```
TTTTCTGGGG GTGTCTAAAT GGGAAGGCGA TAACATGTAG TTGTAATTTA AGTCATAGTG        6000

ATAAATTTGA ATGCGTGTTA CCCATGAGTG ACACATATAA CATGGAGGTG AATCCCTAGA        6060

AATAGGGAAT TAATTGGAAA CTTCGACCAT AATTAGTTTG ATTATATTTA TTCTATTAAT        6120

TGCATTAACC ACTGTATTTG TTGGTTCAGA ATTTGCATTA GTAAAAATTA GAGCAACAAG        6180

AATTGAACAG CTAGCAGATG AAGGAAATAA ACCTGCTAAA ATAGTAAAAA AGATGATTGC        6240

TAATCTAGAT TATTATCTTT CTGCTTGTCA GTTAGGTATA ACAGTAACAT CTTTAGGGTT        6300

AGGTTGGCTT GGTGAACCAA CGTTTGAAAA GCTATTACAC CCAATATTTG AAGCAATCAA        6360

TTTACCAACT GCATTAACGA CGACGATTTC GTTTGCAGTG TCATTTATAA TCGTTACGTA        6420

TTTGCATGTA GTACTTGGTG AATTAGCGCC TAAATCTATA GCTATTCAAC ATACTGAAAA        6480

GCTTGCTTTA GTATATGCAA GACCATTGTT CTATTTCGGT AACATTATGA AACCATTGAT        6540

TTGGCTGATG AATGGTTCTG CACGTGTTAT TATTAGAATG TTTGGTGTAA ATCCTGATGC        6600

CCAAACTGAT GCAATGTCAG AAGAAGAAAT CAAAATTATT ATTAACAATA GTTATAATGG        6660

TGGAGAAATC AACCAAACTG AATTGGCATA TATGCAAAAT ATCTTTTCAT TCGATGAAAG        6720

ACATGCAAAA GATATAATGG TACCTAGAAC TCAAATGATT ACACTAAATG AACCTTTTAA        6780

TGTAGACGAA TTACTAGAAA CAATAAAAGA ACATCAATTT ACGCGTTATC CAATTACTGA        6840

TGATGGTGAT AAAGACCACA TTAAAGGATT TATTAACGTC AAAGAATTTT TAACTGAATA        6900

CGCTTCTGGA AAAACGATTA AAATAGCAAA CTATATaCAT GAGTTGCCAA TGATTTCAGA        6960

GACAACACGT ATCAGTGATG CATTAATTAG AATGCAACGT GAACATGTAC ATATGAGTCT        7020

TATTATAGAT GAATATGGTG GAACGGCAGG TATTTTAACG ATGGAAGATA TTTTAGAAGA        7080

AATCGTTGGA GAAATTCGTG ATGAATTTGA TGATGATGAA GTGAATGATA TCGTTAAAAT        7140

TGATAATAAG ACATTCCAAG TAAATGGCAG AGTACTATTG GATGATTTAA CTGAAGAGTT        7200

CGGTATAGAA TTTGATGACT CTGAGGATAT TGATACGATA GGTGGATGGT TACAATCTCG        7260

TAATACCAAT TTACAAAAAG ATGATTACGT GGATACAACT TATGATCGCT GGGTTGTTTC        7320

AGAAATCGAT AACCACCAAA TTATTTGGGT GATATTAAAC TATGAATTTA ATGAAGCGAG        7380

ACCTACTATC GGACAGTCTG ATGAAGATGA AAAATCAGAA TAGATATTAA TATATAAACC        7440

AACTAAGAAT GATTTAATTC ATTTTTGGTT GGTTATTTTT TTGACTAAAA TTAAnGAAAA        7500

GTGAAAATAG TATTGGAACT CAATATCTTT AATGATTTAA TGAATAAnTT TTATTGAAAG        7560

CGA                                                                     7563
```

(2) INFORMATION FOR SEQ ID NO: 34:

(A) LENGTH: 3492 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```
TTATATCAAC TTCATGGCGG AACCATTGAT GACCCATTAG ACGAAACAAT AAGCGCATTT      60

sATGAATTGA AACAAGAAGG AATTATACGT GCTTACGGTA TTTCTTCTAT TCGCCCAAAT     120

GTAATTGATT ATTATTTAAA ACATAGTCAA ATCGAAACGA TAATGTCTCA ATTCAATTTG     180

ATTGATAATC GTCCAGAATC ATTATTAGAT GCAATTCACA ACAATGATGT TAAAGTATTG     240

GCAAGAGGAC CTGTGTCTAA AGGATTATTA ACTTCAAACA GTGTTAATGT GCTCGACAAT     300

AAATTTAAAG ATGGTATTTT TGATTATTCT CATGATGAAT TGGGTGAAAC AATAGCCTCT     360

ATTAAAGAAA TTGAAAGTAA TTTATCTGCA TTGACATTTA GTTATTTAAC ATCACATGAC     420

GTGCTTGGTT CCATCATTGT AGGTGCAAGT AGCGTCGACC AATTAAAAGA AAATATTGAA     480

AACTATCATA CTAAAGTTAG TTTAGATCAG ATTAAAACAG CAAGAGCTCG TGTAAAGGAT     540

TTGGAATATA CCAATCATTT AGTGTAGAAG TCATTTTCAG TAATAAAAAC AGCAGCATGA     600

GGCGTTTCAT TATAAAAATG CCTTACTGCT GTTGTTTATG TACAATTCGC TATAATTTAT     660

GATTATGATT ACTCACTTAT GATAGAAATT AAAGCGTTGT CCTCACGCAT CAGTATTTAG     720

TAATTTCGCC TTGCGGCATT GCCTTAAGCA AACTTCTGCC ACTTCATCTC TTAATAATTT     780

TATTAAAACA TCTTTCTATA TTTCACTTCG CATGTTGATT CATCATTATT AGTTATTATT     840

TGTACACCCA GCACATTTCC TTGCAACACA AGTAGTTTGA ATTTTTCACA AGTATAATAT     900

AATGTACCGT CTGAAATTTG GTCTACAGAA ATATCGCCTA AAATATCCAG CACTGTAAAT     960

TCTTCAAATA CTGATAGTTG TTCCGCATAT CGTACACAAA GTCTTACCAC ACTCTCCGAT    1020

TGACAGTTCA TTGCCATCCC ACCTATTTAT GCTTTATTTT TAAATAATTT AGGGAAACAT    1080

CGTTCAAAAA ATCTAGGCGC AATTTGATAC ATTTTCAACG CATGaTGCAT CCATTTAGGC    1140

CGATTAATTT CCAATTGTTT TGTTTTAATG CCATAAATGA TATCTTCTGC AAGCTGATTA    1200

GCATCAAGCA TAATTTCCCC CATCTTTTTA gCATACTTCA TTGATGGGTC GGCTTTTTGA    1260

TGAAAAGGTG TATCAATCGG GCCAACATTA ACTGTCATGA TATGTAAGTT TGGTGACTCT    1320

AGTCTTAAAG CATTCATTAA TGCATAAAAC CCTGCTTTCG ATGCCCCATA ATGTGCAGCA    1380

TTTGCTTGTG TGGAAAATGC AGCTTGACTT GAAATACCTA CAATATGTGC GTTAGATGTT    1440

AAATATGGTC TCAACACAGT ATATAAAACA TTAAAACTAA TTAAATTAAG CTGATACGTT    1500
```

```
TAAATGAATC CATCGAATGA TGTATTGTCT TCAAATTGCA GTGCCTGTAT CGACTTCAAA      1620

TCATTTAAGT CACAAGGAAT AACATTTATA GTTTTCCCCA ATTCCTGTTC AAAGATTCTA      1680

GTTGCTTTAT CAACATCACG CACCAACAAC GTTACATGCA CTTTATTTTC TAGTAACTTT      1740

CGGACAATCG ATAAACCTAA ACCACTCGTA CCACCAGTCA CTATAAAATG TTGTCCTTTC      1800

ATCAATTAAC CTTCCTTTTC AATTATATAG AATGCAATTT ATCAACTTTA CATAATTGAG      1860

ACAAGTTGAT TATCTTTCCT AATATATATA CAATAATAAG AAAATATAAC ATACAAATCA      1920

AAAACTAAAG GGATGTGaCG TTAATGrAAC TCGTATTTTA TGGAGCTGGT AATATGGCAC      1980

AAGCTATATT TACAGGrATT ATTAACTCmA GCAACTTAGA TGCCAATGAT ATATATTTAA      2040

CAAATAAATC TAATGAACAA GCTTTAAAAG CATTCGCTGA AAAACTAGGT GTTAACTATA      2100

GTTATGAtGA TGCGACATTA TTAAAAGATG CAGAyTATGT ATTTTTAGGT ACCAAACCAC      2160

ATGACTTTGA TGCTCTAGCA ACACGCATCA AACCACATAT TACAAAAGwC AATTGCTTCA      2220

TTTCAATTAT GGCAGGTATT CCGATTGATT ATATTAAACA ACAATTAGAA TGCCAAAATC      2280

CaGTTGCTAG AATTATGCCA AACACAAATG CGCAAGTTGG ACACTCTGTT ACTGGCATTA      2340

GTTTTTCAAA CAACTTTGAC CCTAAATCTA AAGATGAAAT TAACGATTTA GTTAAAGCAT      2400

TTGGTTCTGT AATTGAAGTA TCAGAAGATC ATTTACATCA AGTAACAGCT ATCACCGGAA      2460

GCGGCCCAGC ATTTTTATAT CATGTATTCG AGCAATATGT TAAAGCTGGT aCsAAACTTG      2520

GTCTAGAAAA AGAACAAGTT GAAGAATCTA TACGCAACCT TATTATAGGT ACAAGTAAGA      2580

TGATTGAACG TTCAGAtTTG AGCATGGCTC AATTAAGAAA AAATATTACC TCTAAAGGTG      2640

GTACGACACA AGCTGGCCTT GATACATTGT CACAATATGA TTTAGTATCT ATTTTCGAAG      2700

ATTGTCTAAA CGCTGCCGTC GACCGTAGTA TTGAACTTTC TAATATAGAA GACCAATAAA      2760

AACAgACCCG CCAACACATG TATGCATCAT CGCAAGCACT GTGTTTGACG GGTTATTTTT      2820

ATAATTTATT GTTATTTGGC AAGCATTGTT TATTACTTTG TCATTAGATT TTAAAACTAT      2880

CAAAATCTTT TACAAAATTA AAATTAGGTG TATCTTCATT TTGTATCAAT GTTTGATAAA      2940

TTTCATTTAT ATCTTCTGTA TTATAGCGAT TGCTCAAATG TGTAATCAAC GTACGTTTAA      3000

CATTGGCTTC TTTTATCAAT GCAAATACGT CTTCAATATG GCTATGATGA TAATTGTTGG      3060

CTAAATGCTT TTCACCATCT ATATAGGTCG CTTCATGTAC CATCACATCA GCATCTCTAG      3120

AAATCACACG TTCATTAGAA CATGGTTTTG TATCACCAAA AATTGCTACA ACTGGACCCT      3180

GTTTGGACTC ACCTCTAAAA TCTTTTGATT GATAAACTTG ACCATTATGT TCAAATGTAT      3240

CATGAGATTT TACTTCTTGA TATTTAGGAC CTGGTTCAAG ACCAATGTTT TTTAACGCTT      3300
```

CATGATTAAG TAAATGCGCC TCTACAGTAA AACCATCCAT GATGATATGT CAGATGATCA   3420

TCGATTTCAA TATATGtAAT TGGATAGTTT AAATGTGACT CTGATAAATT CATAGACATT   3480

TCCACATATG CT   3492

(2) INFORMATION FOR SEQ ID NO: 35:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 1973 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

ATCTAGCGGT ACAAGCGTCT TGGAGGCTAG TATGTTGAAC ATTGTAAACC CTGAAGATCA   60

CTTCGTTGTC ATTGTTTCAG GTGCCTTTGG TAACCGATTT AAACAAATTG CACAAACTTA   120

TTACAAAAAT GTGCATATTT ATGACGTAAC ATGGGGAGAA GCTGTAGATG TCAAAGATTT   180

CATCAATTTC CTTTCAACTT TAAATGTTGA AGTTAAAGCA GTATTTAGTC AATATTGCGA   240

AACATCTACG ACAGTGCTAC ACCCTATTCA CGAGTTAGGA AATGCCATTA ATCAATTTAA   300

TAGTAATATT TATTTTGTAG TTGACGGCGT AAGTtGCATT GGTGCTGTTG ATGTTGACAT   360

TAACAAAGAT AAAATTGATG TACTTGTTTC TGGTAGTCAA AAAGCAATTA TGTTACCTCC   420

AGGATTAGCT TTTGTAGCTT ATAGCCACCG TGCAAAAGAA CATTTCAAAG AAGTAACTAC   480

GCCAAAATTT TATCTAGACT TAAATAAATA CATTTCGTCA CAAGCTGACA ATTCTACACC   540

GTTCACACCA AATGTGTCTT TATTTAGAGG TGTAAATGCA TACGTTGAAA CCGTAAAAGC   600

AGAAGGTTTC AATCACGTAA TAGCACGACA CTATGCAATT AGAAATGCAT TAAGAAGCGC   660

CTTAAAAGCA TTAGATTTAA CTTTATTAGT CAATGATAAA GATGCATCTC CAACGGTTAC   720

AGCATTCAAA CCTAATACAA ATGATGAAGT GAAAATAATC mAAGATGAAC TTAAAAAATnG   780

CTTTAAAATA ACAATTGCnG GTGGTCAAGG CCATCTTAAA GGTCAAATTT TnAGAATTGG   840

TCATATGGGG AAAATTAGTC CTTTCGATAT TTTATCGGTA GTATCTGCTT TAGAAATTAT   900

TTTAACTGAA CACCGTAAAG TTAACTATAT CGGTAAAGGT ATATCAAAAT ATATGGAGGT   960

TATTCATGAA GCAATTTAAT GTACTCGTTG CAGATCCCAT ATCAAAAGAT GGTATCAAAG   1020

CATTATTAGA TCACGAACAA TTCAATGTAG ATATTCAAAC TGGCTTGTCC GAAGAAGCAT   1080

TAATCAAAAT TATACCTTCA TACCATGCTT TAATCGTTCG TAGTCAAACT ACGGTTACTG   1140

AAAATATCAT AAATGCTGCT GATTCTTTAA AAGTAATCGC ACGCGCCGGT GTTGGTGTAG   1200

GTAATACGAT TTCAGCTACT GAACATACAC TGGCAATGTT ATTATCAATG GCACGAAATA   1320

TTCCGCAAGC ACACCAATCA CTTACAAATA AAGAATGGAA TCGAAATGCA TTTAAAGGTA   1380

CTGAGCTTTA TCATAAAACA TTAGGTGTCA TTGGTGCTGG TAGAATTGGT TTAGGTGTTG   1440

CTAAACGTGC GCAAAGTTTC GGAATGAAAA TACTAGCTTT TGACCCTTAC TTAACGGATG   1500

AAAAAGCAAA ATCTTTAAGC ATTACGAAGG CAACAGTTGA TGAGATTGCC CAACATTCTG   1560

ATTTCGTTAC ATTACATACA CCACTAACAC CTAAAACAAA AGGCTTAATT AATGCTGTCT   1620

TTTTTGCCAA AGCAAAACCT AGTTTGCAAA TAATCAATGT GGCACGTGGT GGTATTATTG   1680

ATGAAAAGGC GCTAATAAAA GCATTAGACG AAGGACAAAT TAGTCGGGCA GCTATCGATG   1740

TGTTTGAACA TGAACCTGCA ACTGACTCGC CTCTTGTTGC ACATGATAAA ATTATTGTTA   1800

CACCTCATTT GGGTGCTTCA ACAGTCGAAG CTCAAGAAAA AGTGGCAATT TCTGTTTCAA   1860

ATGAAATCAT CGAAATTTTA ATTGATGGTA CTGTAACGCA TGCAgTGAAT GCACCTAAAA   1920

TGGACTTAAG CAATATAGAT GATACTGTAA AATCATTCAT CAATTTAAGC CAA   1973

(2) INFORMATION FOR SEQ ID NO: 36:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7620 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GGTGTTTCAG ATGTCACTGG TTGATTTTTA ATTGTAGACG GGTATTTTGG GCTTTCGCCA   60

TATTTATTTG CCGGCTTACT GTCAAAGCAT AGGAATACTA TCATAACAAT TGTTAGGCCT   120

AAATAAACAA AATAAAGAAG TACTAACAAA ATATTAAGAC CCATCGGCAT TAATGTAAAA   180

TCACTGTCAT AATAACTATC GATAATCTGT AATACTATAT AAAATATAAT ACTGAATACT   240

GTCATAATCA TTGGAAATAA CATTGTTCTT GATATATCGT GAAATCTTCG AACGCACAAC   300

GCTAAATTTG GAATAAACGT TGCCAAACTA TAGACAAAAG TATACACAGA TGTAAGGATA   360

ATCATCAATA TACTCATAAC TATTAATGTT TCGTTATCCG CCGCTATAGA AATAAAGAAT   420

AGAAATAGGT TTATTATTAG CACACACACA GCTGGAACCA TAAGTATCAA ATGCCATAGT   480

GCCATATACC AATATTCACT ACGTCTTGAT CTCCCCTTAA AATTTACATA ATTTTTCCAA   540

AATAAAACGA ATGATTTCAT AAAACCTACT TGAGGTAATT GTTCCATTGT AATCTCCCTT   600

TCGTTAATCA TATTTATATT TTTAATTATT GTTACCGTTA TAATTTACAA GATTCATTAT   660

```
GTAAAATGAA AACCCGCTAC AAGTACACAT CTATATGGAG ACTCATTTGA AAGTCAACGC      780

TTCGTTAACT ATACTAAAAA TATGTCATAC TGCAATGTTC ACGTTTAAAA GAGTCTCAAT      840

CTATGCAAAT AAAATATTCC ATAACAAAGT ATATACTTTA CATTTTTATA ATTCTTAACA      900

ATACTATTTT ATCAAACATT TACCACAATA AAAATATCTT TTTCATTTTT ATTTAAATTA      960

ATCATATAAT TGCGAGGAGA ATATTATGGA TTTCGTTAAT AATGATACAA GACAAATTGC      1020

TAAAAACTTA TTAGGTGTCA AAGTGATTTA TCAGGATACC ACTCAAACGT ATACAGGCTA      1080

CATCGTGGAA ACGGAAGCTT ACTTAGGTTT GAATGATCGT GCGGCTCATG GCTATGGCGG      1140

TAAAATAACA CCTAAAGTCA CGTCATTATA TAAACGTGGT GGTACAATTT ATGCACATGT      1200

CATGCATACG CATTTACTCA TTAATTTTGT AACAAAATCT GAAGGTATAC CTGAAGGCGT      1260

ACTTATCCGC GCAATTGAAC CAGAAGAAGG TTTATCCGCT ATGTTCCGTA ACAGAGGTAA      1320

GAAAGGCTAC GAGGTAACGA ATGGCCCAGG AAAAATGGACT AAGGCATTTA ACATTCCACG      1380

GGCTATCGAT GGCGCTACGT TAAATGACTG TAGATTGTCT ATTGATACTA AGAATCGTAA      1440

ATATCCTAAA GATATTATTG CTAGTCCACG AATCGGTATT CCAAATAAAG GTGATTGGAC      1500

ACATAAATCT TTACGTTACA CAGTGAAAGG TAATCCATTT GTGTCTCGCA TGCGTAAATC      1560

AGATTGTATG TTTCCCGAAG ATACTTGGAA ATAAATGCCA TCTTTCATTG ATTACTATCA      1620

TGAAAATGAA ATCTATCTCC TTATAAGTCA ATCAATCGTG CCGTCAACAT GCGGATGGGT      1680

TGATTGTTTT TCTTTGTATC CATCATATTT TTTGATTCAT CTCCTCTTAT TGAACTTGTT      1740

CTTAATTATA AAATATAACA ATAGAATTAT TTATAATTAT TAAATTTAGA TGCATTAATA      1800

TTATTGATAT TATTTTCAAA AACTAGAAAT ATTGATTTGT TGCATGTATA ATGTTAAAAG      1860

CGCCCTTTTA TAACGCTTAC ATATAAAGC TTATTTAGGG AGAGGGATAT TCAACAAGGG      1920

GGATTTGAAA ATGATAGAAC TTAATGCAAT TACAACATTA TGTTTAGCTT GTATCCTTTA      1980

TTTACTTGGT AAGGCTATCG TTAATCACGT TAATTTTTTA AAACGTATTT GTATACCAGC      2040

ACCAGTGATT GGCGGCTTAA TCTTTGCTAT TTTAGTTGCG GCTTTGGATT CATTTGGCAT      2100

GGTTAAGATT AAATTAGATG CTTCATTCAT TCAAGATTTC TTCATGTTAG CATTCTTTAC      2160

GACAATCGGT CTTGGTGCAT CATTGAAATT ATTTAAATTA GGTGGCAAAG TCTTGCTATT      2220

ATACTTTATG TTTTGTGCTA TCATTTCAGT CATTCAAAAC ATAGTTGGTG TATCACTAGC      2280

AAAAGTATTA AATATTAAAC CTTTGTTAGG ATTAACAGCA GGTTCCATGT CTATGGAAGG      2340

CGGTCATGGT AATGCTGCTG CTTATGGTAA GACAATTCAA GATTTAGGTA TTGATTCGGC      2400

ACTGACAGCG GCTCTTGCAG CTGCAACTTT AGGTCTTGTA TTTGGAGGGC TTATCGGTGG      2460
```

```
ATTTAAAGAT TATAGCCAAG TAGCATATAA CGAACATTTA CATAGTAAAT TTAATGCCAC    2580

TGAAGTATTC TTCATTCAAT TTACAATCGT TGTATTCTGT ATGGCAGTTG GAAGTTATTT    2640

CAGTCATTTG TTTACAGCTC AAACAGGGAT TAATGTTCCA ATTTACGTTG GCTCATTATT    2700

TGTAGCTGTT ATTGTCCGAA ATATCTCTGA AAGTTTTAAT TTTAATATTG TAGATTTAAA    2760

AATTACTAAT CAAATTGGCG ATGTCGCATT AGGTATTTTC TTATCTCTTG CGCTAATGAG    2820

CATTCAATTA ATCGAAATTT ATAAACTTGC TATACCTCTT ATTATTATCG TTTTAGTTCA    2880

AGTTGTCGTT ATGATTTTAT TTGCTGTTTT AATTTTATTT AGAGGTTTAG GAAAAGATTA    2940

TGATGCTGCA GTAATGGTAG GTGGTTTTAT CGGTCATGGG CTTGGTGCAc GCCAAATGCC    3000

ATGGCAAATT TAGATGTTAT TACTAAAAAA TATGGAAACT CACCTAAAGC ATATTTAGTT    3060

GTACCTATTG TTGGTGCATT CTTAATCGAT TTAATTGGTG TTATAGTCAT TATGGGATTC    3120

ATACAATGGT TTAGTTAAAC ACCAAACTCA TAAATAAAAG AGGAGGCCTT CGCCTCcTcT    3180

TTTATTTATC CTCGATGTAT ATTCAAGTTA CGTTGTTCTA TCCATGACAA TATTTCCGGA    3240

CTAAATACGA TTTGTTTTTG TGTTAAGTCG TCAATATTTT TAGCATCTAA CATCGTCATT    3300

ATTGATTTCA TGTGTTCAAT AAATGATTCT ACATAAGCTA CTGTATGTGC AATGCCATTA    3360

TTTTCAACTT GATTTAAAAA CGGACGTGAC ATACCAGTTG CCTTTGCACC AAGTGCTAAA    3420

CTTTTAATTG CATCGAGTGG TGTACGTAAA CCACCACTCG CGAAAACTGA AATTTCGCTT    3480

TGATAAGCCG TTGTTTCAAG TAATGACTCA ACTGTAGACT GTCCCCATGA TGATAAGTAA    3540

TCCATATCTT TATTTGCACG ACGTTCATTT TCAATATCTA CAAAGTTAGT ACCACCTTTG    3600

CCACTAACAT CGACATACTT GACGCCTATT TGTTGTAAGT CATGCATTAA TTCTTTGCTC    3660

ATACCAAATC CAACTTCTTT TATAATGACT GGAACAGACA CTCGTGATAC AATCGACGCT    3720

ATATTATCTA ACCAAGTCAC AAATTCACGA TTCCCTTCAG GCATAACTAA TTCTTGAGGA    3780

GAATTAACAT GGATTTGTAA CGCTTGTGCC TCAAGTAATT CAACTGCTTC CAAAGCCTTT    3840

TCTACTGGTA CGTCCGCACC AACATTGCTA AAAATCATGC CTTCAGGATT CATTTTTCGC    3900

GCAATCGTAA ACGTCTCAGC CATGCGTGGA TTTCTCAATG CCGCATGTGT TGATCCAACT    3960

GCCATCGCTA AGCCAGTTTC TCTTGCAACT ACAGCTAGCT TTTCATTGAT GTTTTTCGTC    4020

CACTCGCTAC CACCCGTCAT TGCATTAATA TAAACCGGAT ATGCCATCGT TAAGTCAGGC    4080

GTCTGTGATG TCAAATCGAT ATCATTTACA TTAATTGATG GGATAGAATG ATGCACAAAA    4140

CGCATCTTAT CAAAATCTGA ATGCATTGCG TCAGATTGGG CCATTGCTAT TTCAACATGT    4200

TCATTTTTTC TCTGTTCTCT TTGAAAATCA CTCATGATTA AACCTACCTT TTCGTCATTT    4260
```

```
ATTACAGCTA AGCAAATATA ATATCCATAA TGTAAATGTA ATGCCGGCAT ATTTACAAAG     4380

TTCATACCAT AAATCCCAGC TATGAATGTT AACGGTGAAA ATATAACTGA TACTAATGTC     4440

AGTACTTGCA TAATACTATT CATTCTAAAT GACGTGTATG ACTCAAAATT TTCTCGTATT     4500

TCGTTTGTCA TTTCTTGAGC AGTACGAATG ATATTACGTT GCTTAATCAA GTGGTCATCG     4560

ATATGTTGAA TGTATAGCGA ATGTTTATTA TCTATAATCA AATCACCATT TTGTTTCATT     4620

GTATCAATTA GCTCTTGCAT AGGAAACAGT ACACGTTTTA CTTTAATCAA ATCCGAACGT     4680

AACTTAAAGA CACTATCCAT GACCATTTTA TTAAAGCGAT CATCTACATG GCGGTCTTCA     4740

AAATGATAAA CACTATCTTC AAGTGCATAT ACAAAGTTGA AATATTTATC AACCATCATA     4800

TCTAAAATTA ATATGACGAC ATCTGCACAA TCTAATTCTG CATCTAATGT ATTCATATAC     4860

TTATAGACTA CTTTATTTAA TGATTCCAAC GTTTGATGAT GATATGTTAC TAATACATTG     4920

TCTTGTATAA AAATATTTAG TGCTATTGGT GAATAGTTTG ACCCCATAAT ACTATGGAAT     4980

ACTAAGTATT GATAATCTTT ATAAGATTTA TATTTAGCTC GTGGCATACC GTTAATTGCA     5040

TCATCCACTT CTAAATCATT AAAATTAAAA TGTGCTTTAA ACCATTCATT TTCTTGTTCA     5100

TTCGGTTCAT CAAAATCATA CCAAACAATA GTCGCATCTT TTGGTATCTC TTTGATATCA     5160

TCAACTACTT TAAACGGTTC ATATGTAGTT TGATACCGTA TCTTTAAAGC CATCGATACT     5220

CCCCCTAAAT AACGAATTCT CTATTATTTT ATCATGAATT AAATAACGTG TATGTCTTAA     5280

TTTATTTTAG TATGATAGTC ACTAAGGAGA TGGTTATTAT CAAACAACTT TTTACACATA     5340

CTCAAACCGT AACATCTGAA TTCATTGACC ATAACAATCA TATGCATGAT GCAAATTATA     5400

ATATCATTTT TAGTGACGTC GTGAATCGTT TTAATTACAG CCACGGTCTT TCTTTAAAAG     5460

AACGCGAAAA TTTAGCATAT ACGCTATTTA CACTAGAAGA ACATACGACA TACCTCTCAG     5520

AATTGTCTCT TGGCGATGTA TTTACTGTTA CTTTATATAT TTATGATTAC GATTATAAGC     5580

GGTTGCATTT ATTTTTAACA TTAACTAAAG AAGATGGTAC ACTAGCATCA ACAAATGAAG     5640

TAATGATGAT GGGAATTAAT CAGCACACAC GTCGTTCTGA TGCTTTTCCT GAATCATTTT     5700

CAACACAAAT AGCACACTAT TATAAAAATC AATCAACTAT CACTTGGCCT GAACAATTAG     5760

GACATAAAAT AGCAATTCCA CACAAAGGAG CATTAAAATG ACAGATGCAT TACAACAAAA     5820

GATTCATATC GAATTACTAG ATTTATTAGA TGATGTTAAG TTTGAATTAA CAGAATTAAA     5880

TGCACAAAAA GGGTTATACA TTAACGGACC AGCAAATCAG CTACTTAAGC GTGGCGTGCA     5940

TATGGCTTAT GTTCAAGGAC AAAAGCAAGC CATCGATAAT ATTATGACTA TTGTGGAACA     6000

ACAGCTTGAA AGATCAACAT TTCCTAGAAC ATTATGATAA ATTTCAAAAT GAGGTTGCTC     6060
```

```
ATAATTTTTT AGATCAATTT TATCAAATTA AAGGGCAATA CTTTATCATC ACACATATCA      6180

ATACACTTAT TGGTGATTTT CACTCAGAAG CTCATTAACA ATTAGTCTAT ATAACCCTTG      6240

CTATATTTTC AAAAACAAAA CCCAATTACG TTTTCATGTC AAATATCATC TTGCATGAAA      6300

TCGTAACTGG GTCATTTATA TGTTATTAGT TATTTTGTGT TACATCCTCA TCTATCGATT      6360

TGGCAATTTG TTTAATAGCT TTATGTGATT GTCTAATTGG ATAAATTGGA AAATCATGTA      6420

CCATCTTAGG ATAATCATAA AACTCAATGT ATTGATGATG TTGCAACATC ATTTGTTCAA      6480

ATAGCTTCAT ATCAGGATGT GTCATTTCAC GTCCACCACC AAACATATAA ACTGGTGGCA      6540

ATCCTTCTAT TGTGCCATTA ATTGGCGATA TGCGCTTATC TGTTAATGGT AGGCCATTCG      6600

CCCATTTTTT CATAATCTCA TTGACACCAA ACTGACTTAG aACCGCATCT TGTTCGATTA      6660

AGGCGTCCGA AATATCTTTA TTAGATAGTG TTGCATCTAA AATTGGTGAG ATTAAATACA      6720

ATTTATTCGG TAATGGCTGT TGATTAkCTA AAAGAGATTG TACAAAGGAT AATGCCAGTG      6780

CACCACCTGA ACCATCACCC ATGACTACGA CATTTTGATG TCCTACTTCA GATACTAATT      6840

GaTCATAAAC ACGTTGTATC GCTTGGnAAA GTATCGTCaA TATGnAAACT CTGGTGTCTT      6900

TGGATAGATA GGCAGTACAA CCTCATATAA TGtACTTAAA GTGATTTTAT CCCAACAATC      6960

TCCAATGGAA CGGTGATGGT TGTAGTGCAT TGAATCCACC GTGAATATAT AAAATTTTCT      7020

TATCAATTTG ATGTCTGAAA TTAAAGCGAA AGACTTGCAT ATCATCTAAT GACAATTTTT      7080

CTAAATTTGC TTTAACATTT AATGTTGAAG GCTGCTTATG TTTTTTTCTA TTTTCAATTT      7140

CTCTTTTATA AAAAAATCTT TCAACATCTT GATCATTTTT AAACATAATC GAGCGATTGT      7200

GAAGCAAATA TTTATTGACA ACGCTATTCA TAACACGGTT TCTAATCAAT GTCTTAACCT      7260

ACCTTTATAT ATTTTATGTA TCCAATGATk GTCTATCCCC TACATTCTTT GCCAAAAAAA      7320

GTATATAATG TAGAAGATAT TTTCTTTTTC ACTTTCAAAT TTAAGACTAC AATTGAACAG      7380

TGATTTTTCA TCATTATAAC AGACAACTAG ACATATTGAT AAGTAAAGAA AAGAACTTTA      7440

TACGGAGGTA CCTTGCATGA CAAATCCAAA TCAACGATTA GAACCATTTG ATGAGACATT      7500

TCAACAACCG AATATTCATC GTGGTAAGCG ATATGGTAAG AAAAAACGTT CATTGGTAAG      7560

CATGATTATT CAAATCATTG TTGTwATATT AACCACCATC GCTGGAATAC AGCATGGTGG      7620
```

(2) INFORMATION FOR SEQ ID NO: 37:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 9834 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

```
GTCATtACCG amTTTCtTAG AaTCATTTAA AGATGATAAA TATACAAACG TTGGTAATTT        60

AAAAGAAGTG AATTTTGATA AAATTGCTGC GACGAAACCC GAAGTAATCT TTATCTCTGG       120

ACGTACAGCT AATCAAAAGA ATTTAGATGA ATTCAAAAAA GCTGCACCTA AAGCGAAAAT       180

TGTTTATGTT GGTGCAGATG AAAAGAACTT AATTGGTTCA ATGAAACAAA ACACTGAAAA       240

TATCGGAAAA ATTTACGATA AAGAAGATAA AGCTAAAGAA TTAAATAAAG ATTTAGATAA       300

CAAAATTGCT TCAATGAAAG ATAAAACGAA AAACTTCAAT AAAACTGTTA TGTATTTACT       360

AGTTAACGAA GGTGAATTAT CAACATTTGG ACCTAAAGGT CGTTTTGGTG GATTAGTTTA       420

CGATACATTA GGATTCAATG CAGTTGATAA AAAAGTAAGT AATAGCAATC ATGGACAAAA       480

TGTTTCTAAC GAATATGTTA ATAAAGAAAA TCCAGATGTT ATTTTAGCGA TGGATAGAGG       540

TCAAGCGATA AGTGGTAAAT CAACTGCGAA ACAAGCATTA AATAATCCTG TATTAAAAAA       600

TGTTAAAGCA ATTAAAGAAG ACAAAGTATA TAATTTAGAT CCTAAATTAT GGTACTTTGC       660

AGCTGGATCA ACTACAACTA CAATTAAACA AATTGAGGAA CTTGATAAAG TTGTAAAATA       720

ATTTTAAAAG AGGGGAACAA TGGTTAAAGG TCTTAATCAT TGCTCCCCTC TTTTCTTTAA       780

AAAAGGAAAT CTGGGACGTC AATCAATGTC CTAGACTCTA AAATGTTCTG TTGTCAGTCG       840

TTGGTTGAAT GAACATGTAC TTGTAACAAG TTCATTTCAA TACTAGTGGG CTCCAAACAT       900

AGAGAAATTT GATTTTCAAT TTCTACTGAC AATGCAAGTT GGCGGGGCCC AAACATAGAG       960

AATTTCAAAA AGGAATTCTA CAGAAGTGGT GCTTTATCAT GTCTGACCCA CTCCCTATAA      1020

TGTTTTGACT ATGTTGTTTA AATTTCAAAA TAAATATGAT AGTGATATTT ACAGCGATTG      1080

TTAAACCGAG ATTGGCAATT TGGACAACGC TCTACCATCA TATATTCATT GATTGTTAAT      1140

TCGTGTTTGC ATACACCGCA TAAGATTGCT TTTTCGTTAA ATGAAGGCTC AGACCAACGC      1200

TTAATGGCGT GCTTTTCAAA CTCATTATGG CACTTATAGC ATGGATAGTA TTTATTACAA      1260

CATTTAAATT TAATAGCAAT AATATCTTCT TCGGTAAAAT AATGGCGACA scgTGTTTCA      1320

GTATCGATTA ATGAACCATA AACTTTAGGC ATAGACAAAG CTCCTTAACT TACGATTCCT      1380

TTGGATGTTC ACCAATAATG CGAACTTCAC GATTTAATTC AATGCCAAAT TTTTCTTTGA      1440

CGGTCTTTTG TACATAATGA ATAAGGTTTT CATAATCTGT AGCAGTTCCA TTGTCTACAT      1500

TTACCATAAA ACCAGCGTGT TTGGTTGAAA CTTCAACGCC GCCAATACGG TGACCTTGCA      1560

AATTAGAATC TTGTATCAAT TTACCTGCAA AATGACCAGG CGGTCTTTGG AATACACTAC      1620

CACATGAAGG ATACTCTAAA GGTTGTTTAG ATTCTCTACG TTCTGTTAAA TCATCCATTT      1680
```

```
AGTGTTCTTT TTGAATAATG CTATTACGAT AATCTAACTC TAATTCTTTT GTTGTAAGTT    1800

TAATTAACGA GCCTTGTTCG TTTACGCAAA GCGCATAGTC TATACAATCT TTAACTTCGC    1860

CACCATAAGC GCCAGCATTC ATATACACTG CACCACCAAT TGAACCTGGA ATACCACATG    1920

CAAATTCAAG GCCAGTAAGT GCGTAATCAC GAGCAACACG TGAGACATCA ATAATTGCAG    1980

CGCCGCTACC GGCTATTATC GCATCATCAG ATACTTCGAT ATGATCTAGT GATAATAAAC    2040

TAATTACAAT ACCGCGAATA CCACCTTCAC GGATAATAAT ATTTGAGCCA TTTCCTAAAT    2100

ATGTAACAGG AATCTCATTT TGaTAGGCAT ATTTAACAAC TGCTTGTACT TCTTCATTTT    2160

TAGTAGGGGT AATGTAAAAG TCGGCATTAC CACCTGTTTT AGTATAAGTG TATCGTTTTA    2220

AAGGTTCATC AACTTTAATT TTTTCATTTG GGATAAGTTG TTGTAAAGCT TGATAGATGT    2280

CTTTATTTAT CACTTCTCAG TACATCCTTT CTCATGTCTT TAATATCATA TAGTATTATA    2340

CCAATTTTAA AATTCATTTG CGAAAATTGA AAAGAAAGTA TTAGAATTAG TATAATTATA    2400

AAATACGGCA TTATTGTCGT TATAAGTATT TTTTACATAG TTTTTCAAAG TATTGTTGCT    2460

TTTGCATCTC ATATTGTCTA ATTGTTAAGC TATGTTGCAA TATTTGGTGT TTTTTTGTAT    2520

TGAATTGCAA AGCAATATCA TCATTAGTTG ATAAGAGGTA ATCAAGTGCA AGATAAGATT    2580

CAAATGTTTG GGTATTCATT TGAATGATAT GTAGACGCAC CTGTTGTTTT AGTTCATGAA    2640

AATTGTTAAA CTTCGCCATC ATAACTTTCT TAGTATATTT ATGATGCAAA CGATAAAACC    2700

CTACATAATT TAAGCGTTTT TCATCTAAGG ATGTAATATC ATGCAAATTT TCTACACCTA    2760

CTAAAATATC TAAAATTGGC TCTGTTGAAT ATTTAAAATG aTGctACCGC CAATATGTTT    2820

TGTATATTTT ACTGGGCTGT CTAAGAGGTT GAATAATAAT GATTCAATTT CAGTGTATTG    2880

TGATTGAAAA CAATTAGTTA AATCACTATT AATGAATGGT TGAACATTTG AATACATGAT    2940

AAAGTcCTTT GATATTGAAA ATTAATTTAA TCACGATAAA GTCTGGAATA CTATAACATA    3000

ATTCATTTTC ATAATAAACA TGTTTTTGTA TAATGAATCT GTTAAGGAGT GCAATCATGA    3060

AAAAAATTGT TATTATCGCT GTTTTAGCGA TTTTATTTGT AGTAATAAGT GCTTGTGGTA    3120

ATAAAGAAAA AGAGGCACAA CATCAATTTA CTAAGCAATT TAAAGATGTT GAGCAAAAAC    3180

AAAAAGAATT ACAACATGTC ATGGATAATA TACATTTGAA AGAAATTGAT CATCTAAGTA    3240

AAACTGATAC AACTGATAAA AATAGTAAAG AATTTAAGGC ACTACAAGAA GATGTTAAAA    3300

ACCATCTCAT ACCTAAATTT GAAGCATATT ATAAGTCAGC AAAAAATTTG CCTGATGATA    3360

CAATGAAAGT TAAGAAATTA AAAAAAGAAT ATATGACGCT TGCAAATGAG AAGAAGGATG    3420

CGATATATCA ATTAAAAAAA TTCATAGGTT TATGTAATCA ATCTATCAAG TATAACGAAG    3480
```

```
AATTAGCTGA TAATAAAAGT GAAGCAACTA ATCTTACGAC AAAATTAGAA CATAATAATA      3600

AAGCGTTAAG AGATACTGCG AAGAAGAACC TAGATGATAG TAAAGAAAAT GAAGTAAAAG      3660

GCGCGATTAA AAATCACATT ATGCCAATGA TTGAAAAGCA AATTACCGAT ATTAACCAAA      3720

CTAATATTAG TGATAAGCAT GTTAATAATG CAAGGAAAAA CGCAATAGAA ATGTATTACA      3780

GTCTGCAGAA CTATTATAAT ACACGTATTG AAACAATAAA GGTTAGTGAG AAGTTATCAm      3840

AAGTCGATGT AGATAAGTTG CCGAAAAAGG GTATAGATAT AACTCACGGC GATAAAGCCT      3900

TTGAAAAAAA GCTTGAAAAA TTAGAAGAAA AATAACTATA ATCATTTTTC AAAGTTAAAA      3960

ATTTTGAATT TATGGTTAAC ATGTCAACTT ACTATGTGTA TAATGGTAAA CATTGATATT      4020

AACTATATGT ATAAAAATGT CACGCAGATG CTATTTAAAT GTGATAAATA TTTTTAGAGG      4080

TGAATAGAGT GGCTATAAAG CTAAGTTCAA TTGACCAATT TGAACAGGTT ATTGAGGAAA      4140

ATAAATATGT TTTTGTATTA AAACATAGTG AAACTTGTCC AATATCGGCA AATGCGTACG      4200

ATCAATTTAA TAAATTTTTA TATGAACGCG ATATGGACGG TTATTATTTG ATTGTCCAAC      4260

AAGAACGCGA TTTGTCAGAT TATATTGCTA AAAAAACGAA CGTTAAACAT GAATCACCTC      4320

AAGCATTTTA TTTTGTAAAT GGTGAAATGG TTTGGAATCG AGACCACGGT GATATCAATG      4380

TGTCGTCATT AGCACAAGCA GAAGAATAAT GAAACTATAG GGTTGGAACA TTTTGCCTTA      4440

CACTACTAGA CGTGAATAGC ACAACTTAAA TTCGTGTGAA TCAGAGTAGT TTGGCTATAA      4500

TGATGTTCTG ACCTTTTATT TTATGTCACC TTTAGAAGCA GTTAAGTTAG TACTTTTTTA      4560

CAAACATATG TATAATATAT TCGAGTATTT TTATTGAAAa tATTTTGGAA AACGACGAAT      4620

CCAATAAGAA AATTTAAACA TGATTTGTAA GTTAGTTTAA TAGGAAATAT ATGCTAAACC      4680

AAAAGAAGCA TATTGTTATT TACTGGAATA ATTAATAATC ATGTCATGTT AAATGTTAGC      4740

ATATAATCAC GAGATAAAAT CTAAAATTTA AGATTAATCT TTTATGAATA AAAAACGTAT      4800

CACAACAAAT AATAAAGTAA GGTGGTCAAG GTTATGAAAG TATTAGTAGC CATGGATGAG      4860

TTTCATGGAA TTATTTCAAG TTATCAAGCT AATAGATATG TTGAAGAGGC AGTTGCAAGC      4920

CAAATTGAAA CTGCAGATGT AGTTCAAGTA CCATTGTTTA ATGGAAGACA TGAATTATTA      4980

GATTCTGTAT TTTTATGGcm ATCTGGGcaA AAGTATCGTA TACCAGTACA TGATGCAGAT      5040

ATGAATGAAG TTGAAGGTGT TTACGGACAA ACTGATACAG GGATGACCGT TATCGAGGGG      5100

AATTTATTTT TAAAAGGTAA AAAACCAATT GTTGAACGAA CAAGTTATGG TTTAGGAGAA      5160

ATGATTAAAC ATGCATTAGA TAACGACGCA AAACATGTTG TAATTTCACT AGGTGGGATT      5220

GATAGTTTTG ATGCTGGTGC AGGTATGTTA CAAGCATTAG GTGCTCAATT CTATGATGAC      5280
```

```
GATATGTCGA ACTTACACCC TAAAATGGAA ACAGCAAGAA TTCAAGTAAT GTCGGATTTT      5400

TCAAGTCGAT TATATGGTAA GCAAAGTGAA ATCATGCAAA CTTATGATGC GCATCAGTTG      5460

AATCATAATC AAGCAGCAGA AATCGATAAT TTAATTTGGT ATTTTAGTGA GTTATTTAAA      5520

AGTGAATTGA AAATTGCAAT TGGTCCAGTT GAACGTGGTG GTGCTGGTGG TGGAATTGCA      5580

GCAGTCTTGA ATGGACTGTA TCAAGCTGAA ATATTAACCA GTCATGCATT AGTAGACCAA      5640

CTAACACATT TAGAAAATTT AGTTGAACAA GCGGATTTAA TTATTTTTGG AGAAGGATTA      5700

AATGAAAATG ATCAGTTGCT AGAAACGACA ACATTGCGTA TTGCAGAACT TTGTCATAAA      5760

CATCAAAAGG TTGCCATTGC AATTTGTGCA ACTGCTGAAA AGTTTGATTT ATTTGAATCA      5820

CAAGGGGTTA CAGCAATGTT TAATACATTT ATCGATATGC CAGAAACTTA TACTGACTTT      5880

AAAATGGGtT ACAAATTAGG CATTATACGG TTCAGTCTTT AAAACTGTTG AAAACACATT      5940

TTAATGTTGA GGTTTAGTAA AGAAGGACTA AATTGGTGAT GCTGTCATGA TGGTTAATAA      6000

CATTTATGAT GGTTAGCAAA ACGAATTAGA AGATCGAAAG TATACGTAAA AAATATGAAA      6060

AATCACGCTA TCATTGCACT GAATGTTAGC GTGATTTTTA TATATTAATT AAGCCTGAGT      6120

TGAACTAGTA TATAATCGTT GGTTTTTAGT GATTTTCAGC GATATCTTCT ACAATTCCAA      6180

TGATTACTTG TACTGCTTTT TCCaTAACAT CAATGGATGC aTATTCATAT GGGCCGTGGA      6240

AGTTACCGCA ACCTGTAAAG ATGTTTGGAG TTGGTAACCC CATAAATGAC AATTGTGAAC      6300

CATCTGTACC ACCGCGAATA GGTTCAGTGT TTGCTGGAAT ATCTAATTTG GCAAAGACAC      6360

GTTTAGGTAT ATCAATAATA TGAGGCAATG GTAATATTTT TTCTGCCATA TTGAAATATT      6420

GATCCGATAT ATCAACTTTA ACTGGATAAT TTTCAAAATG GGCATTGATA TCGTCACGTA      6480

TTTCTAAAAT ACGTTTCTTA CGCAATTCGA ATTGTTTTTT ATCATGATCA CGAATAATGT      6540

ATTGCAAAGT TGCTTTTTCA ACAGTTCCTT CAAAGTTCAT TAAGTGATAA AAGCCTTCGT      6600

ATCCTTCTGT TCGCTCCGGA ACTTCACTAT CAGGTAGCAA ACTATCGAAT TGTTCACCTA      6660

AACGTATTGC GTTTACCATT GCATTTTTAG CTGAACCAGG ATGAACATTT ACACCGTGGC      6720

ATGTAATAAC CGCTTCAGCA GCGTTAAAGC TTTCATATTG TAATTCTCCA TATTGACTAC      6780

CATCCATAGT ATAAGCAAAA TCAGCATTGA AGCGGTCAAC ATCAAATTTA TGTGGACCAC      6840

GACCGATTTC TTCGTCTGGT GTAAATCCAA TGCGAATGGT ACCATGTTTA ATTTCTGGAT      6900

GTTCTTGTAA ATAACAAATA GCTTCCATAA TTTCCACAAT ACCCGCTTTA TCGTCTGCAC      6960

CTAGTAACGA TGTACCATCA GTTACCATTA ATGTATGACC AACTAAACTG TTAAGTTCTG      7020

GAAATACTTT AGGATCTAAG ACACGTTTAG TATTGCCTAG TTTGTATGGC TTACCATCAT      7080
```

```
GCGCCAAAAA TCCAACTGTT GGGACGTCGA CATCGATGTT ACTTTCTAAT GTAGCAAATA    7200

AGTAGCCATT TTCATCTAAA TCAGTTGGCA ATCCTAATTG TTGTAATTCT TTTTCTAATA    7260

AATGTAACAA ATCCCATTGC TTTTCAGTTG AAGGTGTTGT TGTAGATTTT GGATCAGATT    7320

GCGTATCAAT TGTCGTATAT CTTGTTAATC TATCTATCAA TTGGTTCTTC ATTATATTCG    7380

ACCCCTTAAA CTCTATTATT CATGTTGTAA GATTTTTTAT ATGTCTTACC TTTGATTTTA    7440

CCATACAGTT GTTTGATACG TGTGTATAGG TAATATAGAA TTTCAGAAAC TAATATACCG    7500

AAAGCAATCG CACCTGAAAT CAGTGTAcTT CTAAAAATGT ATTTACAGCA CTTGTATAAT    7560

CATTTGATAC TAAAAAACGA GTCGCTTGAT AAGCTGCACC ACCAGGTACT AATGGTATAA    7620

TGCCTGGCAC TATGAATATA ATTACCGGTC GTTTATATCT GCGACTCATA GTATGACTCA    7680

TTAAGCCTAA AATTAAGCTT CCCAAAAATG AAGCGCCAAC TTTTCCAAAC TCTAAATCTA    7740

CCGTTAATTG GTAAATCGTC CATGCAATGG CACCCACAAA TCCACATGCT ACTAAGAGGC    7800

GTTTGGGTGC ATTGAAAATG ATAGAGAAAA GTACTGTTGA TATAAAGCTG ATTGTAAAAT    7860

GAAATAAATA AAATAGCATG CTTTAACAGT CCTTCCTTAA ATGATTAATA AAACGATTGC    7920

GACACCAGCA CCGATTGCGA ATGCTGTTAA TGCAGCTTCA ACACCGCGAG ACATACCTGC    7980

AAGTAATTCA CCCGCTAATA AATCTCGAAT GGCATTGGTA ATTAATATAC CAGGGACAAG    8040

TGGCATGACA CTGGCTATAG TAATGATATC TTGATTGGTT GCAATGCCTA ATTTAGTAAA    8100

TGTGGCTGCA ATGGATATGA CCACAGCGGC TGCAACAAAC TCTGAGAAAA ATTTAATTTG    8160

TATATAGCGT tGCACAAAGC TGAATGTTAA AAATGCGGAT CCGCCAGCAA TGACTGCAAT    8220

CCAACAATCT GATGCGACAC CACCAAACAT AAATAGGAAG AAGCCACATG CAATGGCAGC    8280

TGCAAAGAAA TTCGTTAAAA AAGAATATTG TAATGATGCA TGCTGTAAAT GAATAAATTC  · 8340

AGATTTAGCT TCATCAATTG TGAGTTCTTT ATTTGATATT TTACGTGAAA GACTATTCGT    8400

TAAAGCGATT TTCTCTAAAT CTGTTGTACG CTCTTGTACA CGAATTAATC TTGTACTTGT    8460

TCGATCGTTT AATGAAAAAA TAATTGCAGT TGAACTGACA AAACTATATG TATTATGAAG    8520

ACCATAACTA TGTGCGATAC GGTTCATTGT ATCTTCAACT CGATATGTTT CAGCACCTGA    8580

TTCaAGTAAA ATTCTACCTG CAATTAATAC AACATCAATC ACTTTGTTTT CATCTATAAT    8640

TGTGATTGAA TCTGGCATAT CAATTCACCT CCAATGATAT GTGTTATTTA TTTGAACAAT    8700

TGaAGTTTAC AACTTGTTGT TACAACTTTC AATAGTGAGA CTTTGTGTTA GTATGATGAA    8760

CTTGTATGGT TCAAATTTAA ATAAGAAAAA CTGTTAATCT TTGCTATTAT ACTATGATTT    8820

AATAATAGCA AAGGATTAAC AGTTTTGTCG TTGTTATAAA TTGATAATAG GGTTAAACAT    8880
```

```
TTTACGCTGT GATTTTGGAT CGTCATCTGT TAAATAACCA ACACCGATAG ACACTGACAA        9000

TTTAATAACT TCTTTGTTTG GTAAATGGAA TGATGATTTT TCAACACCCG AACGAATATT        9060

TTCAGCTAAT TTAACACTTT GATCAAGTGA ATAATTGTGA ATGACAACTG AGAACTCTTC        9120

GCCACCATTT CTAAAAATTT TAAATTGATT CGGCACATAG TTTTTAAGTA ATTGAGACAT        9180

TTGTTTTAAT ACAGCATCAC CTGATTTGTG TGAGTAGGTA TCATTGaCAT CTTTAAATCC        9240

ATCGATATCG ATTAATAATA ATGCGATACT TTGATGTTCT TTTTCAGCTT TTCGTGAAAT        9300

TTCATTTAAA TGTCTATCAA ATTCTTTTAC ATTACCTAAG CCTGTTAAGT AATCATATTT        9360

ATCTTCGTTT TCATAACGAT TTACGAGTGA GAAGAAATGC CAAATATCGA CAAATGTTAT        9420

CGCTGAAGCT AAAGTGATAA TTAATGAAAT TGGTATTAAA ATGATAACTT CCGATAGTGT        9480

GTAAATAGGA CTCACTAACG CGACACCAAA TAAAATGATT ATTGTAACAA CATTAAGTAT        9540

TAATAATGAT AGCACATCAT TTTGTTTTAA AAATGGTCCA ATAGCACTTG TTACTGCAGC        9600

AATAACAATC AACGTAACAC CGTACATAAT CGAGTTGTTA AATACTACAA TTTCAACAAT        9660

TGCTACAATT ACTGTGGCAG ATAATGTATA GACCATATTT GTAAATCTAC CTAAAAACAA        9720

TAAAGGAACG AATGTTAAGT GAATTAAATA ATCTTCACGA TAAGGGATAG GGTAGACAGA        9780

TAATAATAAT GATACGATTG TCATTAAAAC AGTGACATAA GCCTTAGAAA AAAC            9834
```

(2) INFORMATION FOR SEQ ID NO: 38:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 23439 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

```
TCTCAATCAG ATGAAAAATT GCATATCGTA GGTTTTACAG AAAGTGCAAA ATATAATGCG          60

TCATCAGTCA TTTTCACGAA TGACGCTACC ATTGCCAAGA TCAATCCTAG ATTGACTGGA         120

GATAAAATTA ATGCAGTTGT TGTACGTGAT ACAAATTGGA AAGACAAAAA ATTAAACCAA         180

GAGCTTGAAG CGGTAAGTAT TAATGACTTT ATTGAAAATT TACCAGGTTA TAAACCACAG         240

AACTTAACAT TAAACTTTAT GATTTCATTC TTATTTGTCA TTTCAGCTAC AGTTATAGGC         300

ATTTTCCTAT ATGTCATGAC ATTACAAAAG ACGAGTTTAT TTGGCATATT AAAAGCTCAA         360

GGATTTACGA ATGGCTATTT GGCGAATGTG GTAATTTCGC AGACGGTCAT ATTAGCACTA         420

TTTGGTACGG CATTTGGCTT ACTGTTAACA GGCGTTACAG GTGCATTTTT ACCTGATGCA         480
```

```
TCTGTATTAG GAAGTTTATT CTCCATTTTA ACAATTAGAA AAATAGATCC GTTAAAGGCG      600

ATTGGGTAGG AGGTGTAGCA AATGTTGAAA TTTGAAAATG TAACAAAGTC ATTTAAAGAT      660

GGGAATCGTA ACATTGAAGC GGTTAAAGAT ACAAATTTTG AGATAAATAA AGGTGATATT      720

ATAGCATTGG TTGGACCTTC TGGCTCTGGT AAAAGTACAT TTCTAACTAT GGCAGGTGCT      780

TTACAAACAC CGACATCTGG GCACATTTTA ATCAATAACC AAGATATTAC GACAATGAAG      840

CAAAAAGCAT TGGCAAAAGT TAGAATGTCT GAAATAGGTT TTATTTTACA AGCTACAAAC      900

CTTGTACCAT TTTTAACGGT AAAGCAACAA TTTACATTAT TGAAAAAGAA AAATAAGAAT      960

GTTATGTCTA ATGAAGACTA TCAGCAACTT ATGTCACAAT TAGGTCTAAC TTCATTGCTT     1020

AATAAGTTAC CTTCAGAAAT TTCAGGTGGT CAGAAACAAC GTGTGGCGAT AgCaAAGCGT     1080

TATATACGAA TCCGTCGATT ATTTTAGCGG ATGAACCTAC CGCGGCGTTA GATACTGAAA     1140

ATGCGATTGA AGTCATTAAA ATTCTACGTG ATCAAGCCAA ACAAAGAAAG AAAGCATGTA     1200

TTATTGTTAC ACATGATGAA CGACTTAAAG CATATTGTGA TCGTTCATAT CATATGAAAG     1260

ATGGCGTCCT TAATCTTGAA AATGAAACAG TAGAATAGTT TTATTAAGCC GGTACATCAT     1320

GTGCCGGTAT TTTTATGTTT ATGTATTATT TGAATAAACT TTCACATTCA ATTAATAATA     1380

ATTATTATCG AAAATCAGAA ATATTCCGTG AAATATAATA TTTTTTGTAG TAAAATGGCC     1440

TCTAAGTATT CAATATTTAA ATATGGGGAT TGAATATAAA ATTATCGTAA TGGGGGTCAA     1500

TGGTTATGGA TTTATTGATA GGTACTTTAT TTTTATTTTT GGTCTTAGTG ATTTTTACAT     1560

TATTTACATA TAAAGCGCCT AATGGTATGC GTGCCATGGG AGCATTAGCT AATGCAGCAA     1620

TCGCAACATT TTTAGTGGAA GCATTTAATA AATATGTTGG TGGCGAAGTA TTCGGTATTA     1680

AATTTTTAGA AGAGCTAGGA GACGCTGCGG GAGGTCTAGG TGGTGTCGCT GCCGCTGGAT     1740

TAACAGCATT AGCTATCGGT GTGTCACCAG TATATGCATT AGTTATAGCA GCCGCGTGCG     1800

GTGGTATGGA TTTATTACCA GGTTTCTTTG CGGGTTATAT GATTGGATAT GTGATGAAAT     1860

ATACAGAGAA ATATGTGCCG GATGGTGTCG ACTTAATTGG ATCGATTGTC ATCTTAGCGC     1920

CATTAGCTCG TCTTATTGCA GTATTATTAA CGCCAGTAGT GAATAGTACA TTGATTCGAA     1980

TTGGTGATAT TATCCAAAGT AGTACGAATA CGAATCCAAT TATCATGGGT ATCATTTTAG     2040

GTGGTATTAT TACGGTTGTC GGCACAGCGC CATTGAGTTC AATGGCATTG ACAGCATTAT     2100

TAGGTTTAAC GGGTGTACCT ATGGCTATTG GTGCCATGGC AGCATTTAGT TCGGCATTTA     2160

TGAATGGGAC GCTATTCCAT CGCTTAAAAT TAGGTGATCG TAAGTCTACG ATTGCAGTAA     2220

GTATTGAACC TTTATCACAA GCAGATATTG TATCAGCCAA TCCAATTCCA ATCTATATTA     2280
```

```
ATGCGACAGG  TACAGCTACA  CCGATTGCAG  GATTTTTAGT  TATGTTTGGA  TTTAATCATC    2400

CGACGACAAT  TGTGATTTAT  GGTGTAGTAA  TGGCGATTGT  AGGTGCGCTT  GCAGGTTATC    2460

TTGGTTCAAT  TGTATTTAAA  AAATATCCAA  TTGTTACTAA  GCAAGACATG  ATTAATCGAG    2520

GTGCAGTAGA  CGCATAGCAT  CATCATATTG  AATAGTAAAA  ACAAATAAAA  CATAGTAACG    2580

TGATTCAGTC  GATGTAACAG  TCGATAATGA  GTCACGTTTT  TTTATAGAAA  AATACAAGAC    2640

ATAAAAATGT  CATAATTTAT  TGTCGACAAA  TATCATACTG  TATAAACATT  TATCATTTTC    2700

TCAAGTACCT  TTTACACGAT  GGAATGAACT  TACTTTTTAC  GAAATTATGC  GTATTTTATA    2760

AACAAATATC  ATTGATATAA  CGGTAAATGT  AAGCGTTTAC  AACAGAAATA  ACAGCATGCT    2820

ACGATATTTT  TGTAAATTCA  CTGATTCAAG  TATTTTAAGT  CAATATGAGG  AGGGATGTTA    2880

TGAGCGATTC  TGAGAAGAA   ATTTTAAAAA  GAATTAAAGA  TAATCCGTTT  ATTTCACAAC    2940

GTGAACTTGC  TGAGGCAATT  GGATTATCTA  GACCCAGCGT  AGCAAACATT  ATTTCAGGAT    3000

TAATACAAAA  GGAATATGTT  ATGGGAAAGG  CATATGTTTT  AAATGAAGAT  TATCCTATTG    3060

TTTGTATTGG  CGCAGCGAAT  GTAGATCGTA  AGTTTATGT   GCATAAAAAT  TTAGTTGCAG    3120

AAACATCAAA  TCCTGTAACG  TCAACACGCT  CTATTGGTGG  CGTAgCAAGA  AATATTGCTG    3180

AGAACTTAGG  TAGGCTTGGC  GAAACGGTCG  CTTTTTTATC  TGCTAGTGGA  CAAGATAGTG    3240

AATGGGAAAT  GATTAAACGA  TTGTCCACAC  CATTTATGAA  TTTGGATCAT  GTTCAACAAT    3300

TTGAAAATGC  GAGTACAGGT  TCATATACAG  CTTTAATTAG  TAAAGAAGGC  GACATGACAT    3360

ATGGCTTaGC  AGATATGGAA  GTGTTTGACT  ACATTACGCC  TGAATTTTTA  ATTAAGCGTT    3420

CACACTTATT  GAAAAAGGCT  AAGTGCATTA  TTGTAGATTT  GAATTTAGGC  AAAGAGGCAT    3480

TAAACTTCTT  ATGTGCCTAT  ACCACGAAAC  ATCAAATCAA  ATTAGTTATC  ACCACGGTTT    3540

CTTCCCCAAA  AATGAAAAAT  ATGCCTGATT  CATTACATGC  TATTGATTGG  ATTATCACGA    3600

ATAAAGATGA  AACAGAAACA  TACTTAAATT  TAAAAATAGA  ATCTACTGAT  GATTTAAAAA    3660

TAGCTGCTAA  ACGCTGGAAT  GATTTAGGTG  TTAAAAATGT  TATTGTGACA  AATGGCGTGA    3720

AAGAACTCAT  TTATCGAAGT  GGTGAGGAAG  AAATCATTAA  GTCAGTTATG  CCATCAAATA    3780

GTGTGAAAGA  TGTTACAGGT  GCAGGCGATT  CATTCTGTGC  TGCAGTAGTG  TATAGCTGGT    3840

TAAATGGGAT  GTCTACTGAA  GATATATTAA  TTGCTGGTAT  GGTTAACGCA  AAGAAAACGA    3900

TAGAAACGAA  ATATACAGTT  AGGCAAAACC  TAGATCAACA  GCAACTTTAT  CACGATATGG    3960

AGGATTATAA  AAATGGCAAA  TTTACAAAAG  TATATTGAGT  ATTCTCGAGA  AGTTCAGCAA    4020

GCACGGGAGA  ACAATCAACC  GATTGTAGCA  TTAGAATCAA  CAATTATTTC  GCATGGTATG    4080
```

```
GCCATTCCAG CAACCATAGC CATTATAGAT GGCAAAATTA AAATTGGTTT AGAAAGCGAA    4200

GATTTAGAAA TACTGGCAAC TAGTAAAGAC GTTGCTAAAG TATCTAGAAG GGATTTAGCA    4260

GAAGTTATTG CGATGAAGTG TGTTGGTGCT ACTACTGTAG CGACGACGAT GATATGTGCT    4320

GCAATGGCTG GTATTCAATT TTTTGTTACA GGAGGTATTG GGGGCGTCCA TAAAGGTGCA    4380

GAACATACGA TGGACATTTC AGCAGACTTA GAAGAACTGT CTAAAACAAA TGTCACTGTT    4440

ATCTGTGCAG GTGCCAAATC AATTTTAGAC TTACCTAAGA CGATGGAGTA TTTAGAAACA    4500

AAAGGCGTTC CAGTTATTGG ATATCAAACG AATGAATTGC CAGCATTCTT CACTCGCGAA    4560

AGCGGTGTTA AGTTAACAAG TTCGGTTGAA ACGCCAGAAC GACTTGCTGA CATTCATTTA    4620

ACAAAACAGC AGTTAAATCT TGAAGGTGGC ATTGTTGTTG CTAATCCAAT TCCATATGAG    4680

CATGCCTTAT CAAAAGCATA TATTGAGGCA ATCATAAATG AAGCTGTTGT TGAAGCGGAA    4740

AATCAAGGTA TTAAAGGTAA GGACGCCACA CCGTTCTTGT TAGGGAAAAT TGTAGAAAAA    4800

ACGAATGGTA AAAGTTTAGC AGCAAATATA AAACTTGTTG AAAACAATGC GGCGTTGGGT    4860

GCTAAAATTG CTGTCGCTGT TAATAAATTA TTGTAGGTGA TGATACATGA ATATTTTATT    4920

CGCTATCACA GGGATAGCAT TTGCACTATT TGTTGCGTTT TTATTCAGTT TTGATCGTAA    4980

AAAAATAGAC TTCAAAAAGA CGTTAATAAT GATATTTATT CAAGTGTTGA TCGTGTTATT    5040

TATGATGAAC ACAACGATTG GTTTGACAAT TTTAACTGCA CTAGGTTCAT TTTTTGAAGG    5100

GCTAATAAAT ATTAGTAAAG CAGGCATAAA TTTTGTTTTT GGAGATATAC AAAATAAAAA    5160

TGGCTTTACG TTCTTTTTAA ACGTATTACT GCCATTAGTT TTTATTTCTG TATTAATAGG    5220

CATCTTTAAT TATATTAAGG TATTACCATT TATTATCAAA TATGTAGGTA TCGCTATTAA    5280

TAAAATAACT AGAATGGGGC GCTTAGAAAG TTATTTTGCT ATTTCAACAG CAATGTTTGG    5340

GCAACCAGAA GTATATTTAA CAATAAAAGA TATTATTCCA AGATTATCTA GAGCGAAATT     5400

ATATACAATT GCGACGTCTG GTATGAGTGC TGTTAGTATG GCAATGCTAG GTTCATATAT    5460

GCAGATGATT GAACCCAAGT TCGTAGTTAC AGCAGTAATG TTAAATATTT TTAGTGCGCT    5520

TATCATCGCC AGTGTAATCA ATCCCTATAA ATCTGATGAT ACTGATGTTG AAATTGATAA    5580

CTTAACGAAA TCCACAGAAA CTAAAACATT GAATGGAAAA ACAGGAAAAC CTAAGAAAGT    5640

TGCCTTTTTC CAAATGATTG GTGATAGTGC GATGGATGGG TTTAAAATCG CTGTTGTAGT    5700

AGCCGTAATG TTGTTAGCAT TTATTTCATT AATGGAAGCA ATTAATATCA TGTTTGGTAG    5760

TGTTGGTTTG AACTTTAAAC AGCTTATTGG CTATGTGTTT GCACCAATCG CATTCTTAAT    5820

GGGGATTCCA TGGAGCGAAC TGTTCCAGCT GGCTCTTTAA TGGCGACTAA ATTAATTACA    5880
```

```
CAAGGTATCA TTTCAGTTTA CTTAGTAAGc TTCGCTAATT TTGGTACGGT TGGTATCATC    6000

GTAGGTTCAA TTAAAGGCAT TAGTGATAAA CAAGGAGAAA AAGTTGCATC CTTTGCAATG    6060

AGGTTGCTAC TTGGTTCAAC TCTAGCTTCA ATCATTTCAG GATCAATCAT TGGCTTAGTA    6120

TTGTAAATGA ATCGAAGTAC CTAAATTAAA TTCATGGCAA AGCTAAACCC CGTCACCAAG    6180

TTGGCGCAAC AGCGcATgcA TAACTTAGTG ACGGGGTTTT ATCATAACAA TCTACTTTTT    6240

CGTAGCCGTT TTTGAAATGT ATGTTGATGG TTTATCTTTT TCAAAAATTG TTAATCCCGT    6300

TATATCTTTT TTATGTTTTG AAGGGACAAT GAAGCTAAGT ATATAAGCAA AGACAAAAGC    6360

AACTGTAAAT GAAATGGTAG ATACATAGAA AGGTGAGTTA CCTTTGCCAA CACCATTATA    6420

GACATAAGCA AAGATGATAC CCAATATTAA TCCACAAATA ACACCGAATG TATTCGTACG    6480

TTTAGTGAAA ATACCAACTG CAAATACACC AGCCAATGGA ACGCCGAATA ATCCAGTCAC    6540

AAACAAGAAT AAATCCCATA AGTCATTTGA ATTAGAAGCA ATTAAGTATA GTGACATTCC    6600

AAAACCGAAA ATACCTGCAA TGATAATAAT GAAACGTGCA AAGTTAACTT CGTGTCGCTC    6660

GCTACCTTTT CCGAAGAAGC GTTGCTTAAT GTCGATTGAA ATACAAGCAG ATATAGAATT    6720

TAAACTAGAT GAAATGGTAG ACTGTGCAGC GGCGAAAATG GCTGCAATAA GTAATCCTGC    6780

TACAAATGGT GGCATCTCAG TCAAAATGAA ATATGGCACT ACAGATGATG TATTGAAGCC    6840

TTTTGGTAAA ACAGCTTCAT GTGTATAAAA TGAATACAGC ATTGTACCCA TACCATAAAA    6900

TAAGGGTGCT GAAATTAAAG CTAGGATACC ATTTGTCCAT AACGATTTAT TTGTTTCTTT    6960

TAAACTATCA GAAGCTTGAT AACGCTGCAC GACGTCTTGA CTCGCTGTGT ATTGATACAA    7020

GTTGTTGAAA ATATTTCCTA GGAAAATAAT TGGAATGGCA GCTGCCGCAG TATTTAGTTT    7080

CCAATTGTCT GCACTAATTA ATTTTTTGTG CTCAATCGCA TCTGCAAAGA CAGTGCCGAA    7140

ACCGCCTTTA ATGTTCACAA CACCTAGAAT AATAATAACT AAAGCGCCGC CTAATAAAAT    7200

GACGCCTTGA ATGAAATCAC TCCAAACCAC ACCTTCGAAA CCACCTAAAA ATGTATATAA    7260

AATACATAGT AAACCAACGA GTGATGCAAC GATATAAGGG TTCATGTCTG ATACAGATGT    7320

GATTGCTAAT GTTGGTAAGT AGATAACAAT TGCAACACGC CCTAAATGGT AAACGACAAA    7380

TAATAATGAG CCAATGACAC GTATGCTAGG GCCAAATCTA GCTTCTAAAT ATTCATATGC    7440

AGATGTTACC TTTAACTTTT TAAAGAAAGG GACATAGAAA TAAATAAGTA ATGGAATAAT    7500

TGCGACGATA GCAATGTTAC CAGCGATATA TGACCAATCT GTTAAAAATG CTTTCTCTGG    7560

TGTCGACATA AATGTAATCG CACTTAACGT AGTAGCATAA ATTGAAAAGC CAACTACCCA    7620

AGATGGCAAG CGACCACTTG CGGTAAAGAA ACTATTGGTA CTTTGGCTCG CGCGCTTGGT    7680
```

```
TGTGCCAAAT CCAACTTCTT TCATGGGCAA CATCCCCTTT ACAATGTATT GATTCTTTGA      7800

TGTCTATAAA TCGTATTTTG CAATGAGTTG ATCTAATGTT TGTCGATGTG CTTCGTTAAA      7860

AGGTTTGAAA GGTCTTTTCG GTAATCCTGC ATCAATGCCA CGATGACGTA ATATTTCTTT      7920

CAATGTTGGA TAAATCCCCA TTGATAACAC TGTTTCGATA ATGTCGTTTG AATCATGTTG      7980

CAGTTGGTAA GCTTCTTGAA TTTGACCTTG TCGTGCTAAG TCGAAGATTT TTCTTGCACG      8040

GCGACCATTA ACGTTATATG TAGAACCAAT TGCACCATCT ACGCCAGAAA TCGTAGCTTG      8100

AACTAACATT TCATCAAAGC CAGATAAGAT TAATTTGTCT GGGAATGCTT TTCTAATACG      8160

TTCGAGTAGG AAGAAGTTTG GCGCTGTATA TTTAACACCA ACAATTTTTT CATGATTAAA      8220

TAGCTCGCTG AATTGTTCAA TAGAAATATT CACACCTGTT AAATCTGGTA TTGCATAAAT      8280

AATCATATTG TTCTGAGTTG CTTCGATAAT ATCGAAATAG TAATCTCTAA TTTCTTCAAA      8340

AGTAAATGGA TAGTAGAATG GTGTTACGGC AGAAAGTGCA TCATAACCGA GTTCTGTGGC      8400

ATATTTTCCA AGTTCAATGG CTTCATTTAA ATCTAACGAA CCTACTTGAG CAATCAATTT      8460

CACTTTATCC CCAACTGCCT CTTTGGCAAC CTTGAAAACT TGCTTCTTCT GCTCTGTATT      8520

TAATAAAAAG TTTTCGCCTG AGCTACCATT TACATAAAGA CCGTCTAATT CTTCAGTTTC      8580

AATGGCATTT TGAGCAATTT GTTTAAGTCC TTGTTCATTT ACTTGACCAT TTTCATCAAA      8640

AGGAACGAGT AACGCTGCAT ATAAACCTTT TAAATCTTTG TTCATTATGA AGTCCCTCCA      8700

AAAATCATTT GATAATATAG TTTACAGCTA TAATTGTAAA CGCTATCATA AAATGTAACA      8760

ATATCTTTTT GAAAATTGTA GTCATATTTA TGTATAATTA ATGAAAATGT TTTTCAAAAT      8820

CAATAGAAAT GGAGTGAGTA AGGTGTATTA CATCGCAATC GATATTGGAG GCACTCAAAT      8880

TAAATCGGCA GTTATTGATA AGCAATTGAA TATGTTTGAC TATCAACAAA TATCAACGCC      8940

GGACAACAAA AGTGAGCTTA TTACTGACAA AGTATATGAG ATTGTAACAG GATATATGAA      9000

GCAATATCAG TTGATCCAAC CTGTCATAGG TATTTCATCA GCAGGCGTTG TTGATGAACA      9060

AAAAGGCGAA ATTGTATACG CAGGGCCAAC CATTCCGAAT TATAAAGGTA CTAATTTTAA      9120

GCGATTATTA AAATCACTGT CTCCTTATGT CAAAGTAAAA AATGATGTAA ACGCTGCATT      9180

ACTAGGCGAA TTGAAATTAC ATCAATATCA AGCAGAACGG ATCTTTTGTA TGACGCTTGG      9240

TACAGGCATT GGGGGTGCGT ACAAGAATAA TCAAGGTCAT ATTGATAATG GTGAGCTTCA      9300

TAAGGCAAAT GAAGTTGGGT ATTTATTGTA TCGTCCAACT GAAAATACAA CGTTTGAGCA      9360

ACGTGCTGCA ACGAGTGCAT TGAAAAAGCG CATGATTGCC GGAGGATTTA CGAGAAGCAC      9420

ACATGTGCCA GTATTGTTTG AAGCAGCTGA AGAAGGTGAT GATATTGCAA AACAAATATT      9480
```

```
AGGGCTTATA TTAATTGGGG GCGGTATATC TGAACAAGGA GATAATCTCA TTAAATATAT      9600

CGAGCCGAAA GTTGCACACT ATTTACCAAA AGACTATGTT TATGCACCAA TACAAACGAC      9660

TAAGAGTAAA AATGATGCAG CATTATATGG CTGTTTGCAA TGATAGTTGA AAGAAGGAGT      9720

CATTCTAAAA TAGAATTTGA AACCGTTACG AGAGATGAGA GCTGTTGTTA GTTCCACACA      9780

TCACACTCTA TCTAGGACCA ATCTAAACTA TATCAACCAA CAGTGTGCCA CGGGCAAATT      9840

AAATTGAAGA AGCTGAGATA TTAAAATTTT AGAAATGTA AAAAAATATT TGGTATTGAA      9900

ATTAAAAAAG CACCTAGCAA CTCGTTGGGA CAATCACGAT GATTGTCTAC AGTTGCAGGT      9960

GGATTTGAAT ATACTACTAG TTATTTGTTG TCTAGGATAA TAGATTTAGT ATGTTGATAA     10020

GTTTGACTCA GATTCGTATT TTCTAATAAA TGATAACTCA CGATATCGAT TAAAAAGAGT     10080

GTCGCAATTT GTGTGTTGAT AAATTGATGG TCGGTATTAC GCGATTGATC CGTTGTTAAA     10140

AGTACTAAAT CTGCACAATC TGTAAGTTTA CTACCTTCAA AATTTGTGAT GGCAACGACA     10200

TATGCACCAT GAGATTTGGC GACTTCCGCT GCAGAAATTA ATTCCGAAGT ATTACCACTA     10260

TTTGACATAG CAATAAACAT ATCCGAATGA GATAGTAGGG ATGCCGATAT TTTCATTAAA     10320

TGTGAATCGG TAGTAACATT ACCTTTTAGC CCCATACGAA TCATACGATA ATAAAATTCA     10380

GTCGCTGATA AACCAGAGCT ACCTAGTCCA GCAAAGAGTA TATGTCGACT TGATTGAAGT     10440

TTGTCGATAA AGGTTTGGAT AATGTCGTTA TCAATAAATT CACCAGTTTG TTGAATGATT     10500

TGTTGATGAT ATTTATGAAT TCTTTGAATA ATTGGGCTAT TTTCAATAAC TGTCTCTGTC     10560

ATTTCTTGTT GAATATTAAA TTTTAAATCT TGGAAATTCT CATAATCCAG CTTATGACTA     10620

AAGCGTGTCA TCGTTGCTGG TGATGTACCA ATCGCATGGG CTAAGGAGTT AATCGTTGAA     10680

AAGGCATCGC TATAACCATT TTGTCTTATA TAATTGACGA TGCCGTTTATC AGTTTTTGTA     10740

AATAAATGTT GATAACGTTG AACACGATTC TCAAATTTCA TTGTGTCACC CCTTCATCTT     10800

AATGATTACT ATTATATATG AAAAATATTT TCAAGATAGT AAAAAGCATT GATAAAAATT     10860

ATCTTAATGA TATATTGTAA ATGACTTTAC GTGAAAAAAC GACTTATGGA GTGAGGAATA     10920

ATGTTACCAC ATGGATTAAT AGTATCTTGT CAGGCACTAC CAGATGAACC ATTGCATTCA     10980

TCTTTTATTA TGTCGAAAAT GGCATTAGCT GCGTATGAAG GTGGTGCTGT TGGTATTCGC     11040

GCAAATACTA AGGAAGACAT TTTAGCAATT AAAGAAACGG TAGATTTACC AGTTATTGGC     11100

ATTGTGAAAC GTGACTATGA TCACTCAGAT GTTTTCATTA CTGCAACGTC AAAAGAAGTT     11160

GATGAACTGA TAGAAAGCCA ATGTGAAGTC ATTGCATTGG ATGCAACGTT ACAGCAACGT     11220

CCGAAAGAAA CGTTAGACGA ATTAGTATCA TATATTAGAA CACATGCACC GAACGTTGAA     11280
```

```
TATATTGGCA CGACGTTACA TGGCTATACT AGTTATACGC AAGGACAATT ACTTTATCAA      11400

AATGACTTCC AATTTTTAAA AGATGTACTA CAAAGTGTTG ATGCAAAAGT TATTGCGGAA      11460

GGTAATGTCA TTACACCGGA TATGTATAAA CGTGTGATGG ACTTAGGCGT TCATTGTTCA      11520

GTCGTTGGTG GTGCGATAAC ACGACCAAAA GAAATTACGA AACGTTTTGT TCAAATTATG      11580

GAAGATTAAA TGATAACGAT AAAAAAACGA GATGACCATC ATTAATTAAA GGCACCTAAT      11640

TATCTTAGGT GGCTGAATGA ATGTAATGGG TTCATCTCGT TTTGTTTGTT TATGATAGTG      11700

ATTTTATTTT CAACTTTATC CAAAAATAAG TAAAGCGACG GGGATGGTGA TTAATAGCGA      11760

CAACGCCACG CGTAAAAACC AAATGATGAT GAGTTTCCAG ACAGGTATTT TAATTTCAGT      11820

TGCTAGTATA CATGGCACTA ATGCTGAGAA AAAGATAATG GCTGATACGC TTACTACACC      11880

GACGACAAAT TTAGTACTCA TTGCAGCTTT AGTTACTAAC AAAGATGGTA GAAACATCTC      11940

TACAATAGAA AckCTGACGC TTTTGCTAGT AAAGCCTGAT CAGCAATTGG GAAAATATAA      12000

ATAAATGGAT AGAAGATATA GCCAAGCCAA TCAATGAATG GTGTATAGTT CGCTACAATC      12060

AGTCCTAAAA AACCAATCGA TAATATAGAA GGTAAAATAC CAACAGTCAT TTCTAAACCG      12120

TCTTTCAAAT TGTCCCAAAC GTTCTTCACG AGAGATGGTG TTAATGCATT TTGTTTCATC      12180

GCCTCTGCAT ATGCAGTTTT CAGTCTGCTT CCTTCAATAG CAACTTCTTG TTCTCCTTCT      12240

TGTCCGTTAT AATATTCTGT TGATTCATTG CTGATTGGCG GTAGCCATGC AGTAATTGCA      12300

GTCACGACAA ATGTGATGAC TAAAGTTATC CAAAAGTATA AATTCCAATG CGGCATTAAT      12360

CCTAAAGTTT TAGCAACGAT AATCATAAAA GTTGCTGAAA CTGTTGAAAA GCCAGTCGCA      12420

ATAATCGTGG CTTCTCGTTT GTTGTACATC CCTTGCTTAT AGACACGATT AGTAATCAAT      12480

AATCCTAAGG AATAACTGCC GACAAACGAA GCCACTGCAT CGACAGCGGA TTTTCCTGGT      12540

GTTTTAAAAA TAGGTCTCAT AATAGGCTCC ATATAAACAC CGACAAATTC TAATAAGCCA      12600

TAGCCCACTA ATAAAGAAAG cGcAATTGCA CCTACTGGAA TTAAGATACT TAATGGCATC      12660

ATTAATTTTT CAAACAAAAA CGGACCATAG TTAGCTTTAA ATAGTATTGA TGGACCGATT      12720

TTAAATACAT ACATTATACC GATCATTGCA CCTGCAACTT TAAATAATGT AATGACCAAG      12780

TTTGTGATTG AAGTCATAAA AGTACGTCTC ACTATTGGTA ACGCTGTACC AATTAAAATC      12840

ATAATCAGTG CAACATAGGG CATAAGTGGA CCTATGATTG AGCGAATGGC TAGATGAACA      12900

TGATCGACGA AAATAGTGTT GTTACCATTA ATCGTAAAAG GAATAAAGAA ACATAGTATG      12960

CCCACTAAAC TATAGACAAA AAAACGCCAT GCACTTGGTT GTTGTGCATT AGAATGATAT      13020

TGATTCATTA AAGCAACCCC TTTGTTTAAA TGAATACACA AAACTGTATG ATGCATCTTC      13080
```

```
      ATAGTTTGAA TTATTTTCAT ACCAATACAA ATTAACTAAT TATATATAGA TTGAAACTAT    13200

      ATTACTTAAT AAAATATTTA TCTTAAATGT TGTTGTGTTG ATTCAACACC ACAACTAAAA    13260

      GTGTTTATAA ATTATTTGGA AATACACATA TTTGTAAATG ATTAGTATCG ATTTAATATC    13320

      GTATTATTAA ATTTTTATTA ATTTTGTAGT CTTAATCmAA AAATAATATA TGTCATGTTA    13380

      TATTGAAGGT GCAGTTGTTT TTCATTCTCA AGAGGGGGTC AAAAAAATAC TTTTGAGGTG    13440

      ATTATATGTT AAGAGGACAA GAAGAAAGAA AGTATAGTAT TAGAAAGTAT TCAATAGGCG    13500

      TGGTGTCAGT GTTAGCGGCT ACAATGTTTG TTGTGTCATC ACATGAAGCA CAAGCCTCGG    13560

      AAAAAACATC AACTAATGCA GCGGCACAAA AAGAAACACT AAATCAACCG GGAGAACAAG    13620

      GGAATGCGAT AACGTCACAT CAAATGCAGT CAGGAAAGCA ATTAGACGAT ATGCATAAAG    13680

      AGAATGGTAA AAGTGGAACA GTGACAGAAG GTAAAGATAC GCTTCAATCA TCGAAGCATC    13740

      AATCAACACA AAATAGTAAA ACAATCAGAA CGCAAAATGA TAATCAAGTA AAGCAAGATT    13800

      CTGAACGACA AGGTTCTAAA CAGTCACACC AAAATAATGC GACTAATAAT ACTGAACGTC    13860

      AAAATGATCA GGTTCAAAAT ACCCATCATG CTGAACGTAA TGGATCACAA TCGACAACGT    13920

      CACAATCGAA TGATGTTGAT AAATCACAAC CATCCATTCC GGCACAAAAG GTAATACCCA    13980

      ATCATGATAA AGCAGCACCA ACTTCAACTA CACCCCCGTC TAATGATAAA ACTGCACCTA    14040

      AATCAACAAA AGCACAAGAT GCAACCACGG ACAAACATCC AAATCAACAA GATACACATC    14100

      AACCTGCGCA TCAAATCATA GATGCAAAGC AAGATGATAC TGTTCGCCAA AGTGAACAGA    14160

      AACCACAAGT TGGCGATTTA AGTAAACATA TCGATGGTCA AAATTCCCCA GAGAAACCGA    14220

      CAGATAAAAA TACTGATaAT AAACAACTAA TCAAAGATGC GCTTCAAGCG CCTAAAACAC    14280

      GTTCGACTAC AAATGCAGCA GCAGATGCTA AAAAGGTTCG ACCACTTAAA GCGAATCAAG    14340

      TACAACCACT TAACAAATAT CCAGTTGTTT TTGTACATGG ATTTTTAGGA TTAGTAGGCG    14400

      ATAATGCACC TGCTTTATAT CCAAATTATT GGGGTGGAAA TAAATTTAAA GTTATCGAAG    14460

      AATTGAGAAA GCAAGGCTAT AATGTACATC AAGCAAGTGT AAGTGCATTT GGTAGTAACT    14520

      ATGATCGCGC TGTAGAACTT TATTATTACA TTAAAGGTGG TCGCGTAGAT TATGGCGCAG    14580

      CACATGCAGC TAAATACGGA CATGAGCGCT ATGGTAAGAC TTATAAAGGA ATCATGCCTA    14640

      ATTGGGAACC TGGTAAAAAG GTACATCTTG TAGGGCATAG TATGGGTGGT CAAACAATTC    14700

      GTTTAATGGA AGAGTTTTTA AGAAATGGTA ACAAAGAAGA AATTGCCTAT CATAAAGCGC    14760

      ATGGTGGAGA AATATCACCA TTATTCACTG GTGGTCATAA CAATATGGTT GCATCAATCA    14820

      CAACATTAGC AACACCACAT AATGGTTCAC AAGCAGCTGA TAAGTTTGGA AATACAGAAG    14880
```

```
ATTTAGGATT AACGCAATGG GGCTTTAAAC AATTACCAAA TGAGAGTTAC ATTGACTATA      15000

TAAAACGCGT TAGTAAAAGC AAAATTTGGA CATCAGACGA CAATGCTGCC TATGATTTAA      15060

CGTTAGATGG CTCTGCAAAA TTGAACAACA TGACAAGTAT GAATCCTAAT ATTACGTATA      15120

CGACTTATAC AGGTGTATCA TCTCATACTG GTCCATTAGG TTATGAAAAT CCTGATTTAG      15180

GTACATTTTT CTTAATGGCT ACAACGAGTA GAATTATTGG TCATGATGCA AGAGAAGAAT      15240

GGCGTAAAAA TGATGGTGTC GTACCAGTGA TTTCGTCATT ACATCCGTCC AATCAACCAT      15300

TTGTTAATGT TACGAATGAT GAACCTGCCA CACGCAGAGG TATCTGGCAA GTTAAACCAA      15360

TCATACAAGG ATGGGATCAT GTCGATTTTA TCGGTGTGGA CTTCCTGGAT TTCAAACGTA      15420

AAGGTGCAGA ACTTGCCAAC TTCTATACAG GTATTATAAA TGACTTGTTG CGTGTTGAAG      15480

CGACTGAAAG TAAAGGAACA CAATTGAAAG CAAGTTAAAT TCATCTTCTG AATTTAATAT      15540

GCTATGTAAA TCGTGCTGTT ATCATGGCAC ATCAGATATA AGTAGCATCA CAGTGTTGAA      15600

TTTAAAAATA GTAAAGTGAA ATAAAGCGCC TGTCTCATTA GCGAAAACTA AAGGGACAGG      15660

CGTATCTGTT TATGAGCTTA ATAAATTGTA TGAATAATAT GGTTGATCGA ATAACTGTTT      15720

ATCATGATGA TAAATTGAGT TTTTTAAAAT AATGATATAT TACATCATTG TTATAGCGTT      15780

TAAGAAATCA ACAACTTTAC GATAAATAGT GATTGCTTCG TCATTAGGTC TACGATCAAA      15840

ATCATGCTCG TTTTTATTCA CGCGTTCAAA TGTTGAATGT GGAACATGAT TCATGATATG      15900

TTCGCTTTCC TCAACGGGAA CATCATAATC GCCATTACAA TGCGCAATGA AAACAGGTGG      15960

AAGTGTTTTA AGTTCATCTG GTGCAATATT ATATTTGAA TTAGTATAAT CAGCAATGTT      16020

AATCATATTT ATCCATTTAC CTGTGCCACG TGCATAAACG TAGATTAAAA AACGTTGTGC      16080

GATTTGATCT TGAACAACCG GTGTTGGTGA AGTGAGTTGT GCAATCATTG TTTCGTTTAC      16140

GCTTTGAGCT ATTTTTGCGT AATAACTATT AGTTGTTTTA AAAGGTTCAG TGTTGATGCG      16200

ACTATAACCA TAAAAATCAA TAACACCATC AATATCTCTG TCTCGTGCAA TTAATAGACT      16260

TAAATATGCA CCTGATGATC TGCCAAAGGT AAAAATAGGG CAATTAGAAT ATTGTGATTG      16320

AATCGCATCG AATGAtGCgn AGnACATCCT CAATAATGCA ATCGAGACTT ACTTCTGGTA      16380

ATAAACGATA ACTTAGTTGA ATTAAATCGT AATGTTCCGT AAgATATCGA TATACTGTGG      16440

GGATAAATCG TTAGCTTTAC CGAACATTAA TCCACCACCG TGGATGTAGA CAATAGCGCC      16500

TTTTGTTGGT TGATTTTTTG CTTTAATAAT TGTGTAAGGT AATGCAAATG CATCTTTAGT      16560

AATTACTTTA TCTTTAATTT CAGTCACGAT TTAATAGGCT CCTTATTTTT GATATTGATG      16620

TCATTATAAC ACTGTCTTAA ATTTCCATGA AAAATAGTCT TAAGACGATG AGTCATGATA      16680
```

```
CATCATTTTA ACAATATCTT TAAAAGCAGC ATGTGGAATG GCTAAATCTT CTAAATCTGC    16800

CATAGAAAAT TCAAGATTGA TATCATGTGG TCGCTGTTCA GCAAGTTTAT GCACAAAGTC    16860

AGGTTCTGTG ACAAAAGGCG AAGACATGCC GACCATATCT GCATGTTGTA AAGCATCTAA    16920

AGCAGACTCT GGAGAATTAA TCCCGCCACT TGCAATTAAA GGGATACGAC CTGCTAAATG    16980

TTCATAGACA ATTTGGTTAA CTGGTCGACC GAAATGATCA CCTGGTGTAC GAGACGTATT    17040

TTGATAAATA TGTCGACCCC AGCTAGCGAT TGCTAAGTAT TGGATGTTTG AAACGTCCAT    17100

GACCCAATTG ATTAATTGGT TGAACTCGTC AATGGTATAT CCTAAATCAC TGCCTCTGGT    17160

TTCTTCTGGC GTTGCTCGAA ATCCTAAAAT AAAATTGTCA GGTGCTTCTT TATCAATCAC    17220

TTCTTGTACC GCACGCATAA CTTCTAAACA TAATCTTGCA CGATTTTTTA ATGAGTCGGC    17280

ACCGTAATGG TCTGTACGTT TATTCGAAAA AGTTGAGAAA AATGTTTGAA TCAGCAAACG    17340

TTGTGCAATC GAAATTTCCA CACCATCAAA ACCTGCTTTA ATCGCGCGTA ATGTAGCATC    17400

GCGATACTGC TGAATGATGC TATTGATTTT CTCATGAGAC ATGGCGATAA CATCGTGTTC    17460

AATCGGTGAA TGCAATGTCA TAGGGCTTGG TCCATACACC TTTCCAAAAT TTAAAATGGC    17520

TTGATTTGAA AAACGACCAG CATGCGCTAg CTGGATAATA GCGAGGCTAC CATGTTGTTT    17580

CATCGTAGAT GCCATGTTAG TTAATCCAGG GATACAAGCA TCATGATCAA TATTAAAGCC    17640

ATATTCAAAC AATTGACCAT AAGGTTCAAT GTAAGCAGCG CCGGTGACTT GCATTCCAGC    17700

TGAATTAGAG CGACGTGCAG CATAAGCCAA GTCTTCTTTT GTAATATAGC CTTCTTTTGT    17760

TGATGTGTTT ACGGTCATTG GTGATAATAC AAAGCGATTC GAAATTTTGA TGCCATTAGG    17820

TAAGTGGATT GATTGTAAAA GTGGTTTGTA TCGGTACATA CTATGATTCC TTTTCTATTC    17880

AATATTGTTT TCAAAGTACC ATGGAAAGAA TGAATAATCA ATGATGAACA GTCTTGATAG    17940

AATAGAATTG GTACATGGAA AGTATTTTTA AAATTAAACT AATGAATGGC ATTTGTAGGT    18000

CTGAAAATAT GAATATGAAA AAGAAAAATA AAGGCGAAAA GATATAAAAG TTAATTGAAA    18060

AACGTTATCA TATACGTGGG TATATGAAGA GGGAATGGTA TTAAGAACGC TAAAATGTTA    18120

TGTCGGTTTG ACATGACAGG ATAAGTTTGG AGATGACGGA TTGGTTAAAT TAAGCGTATT    18180

AGACTATGCC TTAATAGATG AAGGTAAGGA TGCACAAAAG GCATTGCAAG ATTCAGTGAC    18240

ACTTGCAAAA TTAGCAGATC GACTTGGCTT TAAGCGAATT TGGTTTACGG AACATCATAA    18300

TGTACCAGCG TTTGCGTGTA GTAGTCCAGA ACTTTTGATG ATGCATACAT TGGCGCAGAC    18360

AAATCACATA CGAGTTGGCT CTGGTGGTGT GATGCTGCCG CACTATCGAC CTTATAAAAT    18420

TGCTGAGCAT TTTAGAATGA TGGCAGCGTT ATATCCAAAT CGTATTGATT TAGGTATTGG    18480
```

```
        TAGTTACGAT GAATCGATTT CGTTATTACG TGATTATCTT ACAATAAAGG ATAAACCAAG      18600

        TGCGCATACG TTAGGTGTCC AACCACACAT TGATCATTTT CCAGAAATGT GGTTATTAAG      18660

  5     TAGTAGCGCA ACATCTGCCA AAATAGCTGC CGAACTAGGT ATAGGGCTTT CTGTTGGAAC      18720

        ATTTTTGCTA CCAGATATAA ATGCGATACA TACAGCGAAG GATAACATTG ATATTTACAA      18780

        AAAACATTTC CAAGCATCAA CGATTAAAAT GGACGCAAAG GTGATGGCAT CTGTATTTGT      18840

  10    CATTGTAGCT GATAACGAAG CGGAAGTAGC AGCATTACAA CATGCCTTAG ATGTTTGGTT      18900

        ATTAGGTAAA TTACAATTTG CAGAATTTGA AGATTTTCCT TCAGTAGACA CAGCACAAAA      18960

  15    GTATAAGCTT AATGATCGAG ACAAAGAGAT GATTCAAGCA CATCAAGCAC GCATCATTGC      19020

        AGGTACACAA GAAAAGGTTA AAGCACAATT AGATGATTTC ATTGCTACGT TTGAAGTTGA      19080

        TGAGGTGTTA GTAGCACCGC TTATTCCAGG TATTGAACAG CGTTGTAAAA CATTAAAATT      19140

  20    ACTCGCGGAA ATTTATTTGT AGCATTTTAA ATAGAAGAGA AAGGATGAAG ATAAGATGAA      19200

        AAAGTTAGCC AATTATTTAT GGGTAGAAAA AGTAGGAGAT TTGTATGTGT TTAGTATGAC      19260

        ACCTGAATTG CAAGATGATA TTGGGACAGT AGGTTATGTT GAATTCGTAA GTCCAGATGA      19320

  25    AGTTAAAGTG GATGATGAAA TTGTGAGTAT CGAAGCATCG AAAACGGTCA TTGATGTGCA      19380

        AACGCCATTG TCAGGAACGA TTATTGAGCG AAATACAAAA GCGGAAGAAG AACCGACAAT      19440

        TTTAAACTCT GAAAAACCAG AAGAAAATTG GTTGTTCAAA TTGGATGATG TCGATAAAGA      19500

  30    AGCATTCCTA GCATTACCGG AGGCTTAAAT GGAAACGTTA AAATCAAATA AAGCGAGACT      19560

        TGAATATTTA ATCAATGATA TGCATCGAGA GAGAAATGAC AATGACGTAT TGGTAATGCC      19620

        ATCTTCATTT GAAGATTTGT GGGAATTATA TCGAGGCTTA GCAAATGTCA GACCGGCATT      19680

  35    ACCTGTAAGT GATGAATATT TAGCTGTACA AGATGCTATG TTAAGTGATT TGAATCGTCA      19740

        ACATGTTACG GATTTGAAGG ATTTGAAGCC GATAAAAGGT GACAATATCT TTGTTTGGCA      19800

        AGGTGATATC ACGACGTTAA AAATCGATGC TATTGTTAAT GCTGCAAATA GTCGTTTTCT      19860

  40    AGGATGTATG CAAGCTAATC ATGACTGCAT TGATAATATT ATTCATACAA AAGCGGGTGT      19920

        TCAAGTTCGA CTTGATTGTG CAGAGATCAT TCGACAACAA GGGCGCAATG AAGGTGTAGG      19980

  45    TAAAGCCAAA ATAACACGTG GATATAATTT GCCAGCAAAG TATATAATTC ATACGGTTGG      20040

        TCCGCAAATA CGTCGATTGC CTGTTTCAAA GATGAATCAG GACTTGTTAG CTAAATGTTA      20100

        TCTTAGCTGT CTTAAATTGG CTGATCAACA TAGTTTAAAT CATGTCGCTT TTTGCTGTAT      20160

  50    ATCTACAGGT GTATTTGCTT TTCCTCAAGA TGAAGCAGCA GAAATTGCTG TTCGAACAGT      20220

        AGAAAGCTAT CTCAAAGAAA CAAATTCAAC ATTGAAAGTC GTGTTCAATG TATTTACAGA      20280
```

55

```
CAATGTCTCT GTTAATGGAT GACAAGACAA AGCAGGCTGA AGTATTGCGT ACTGCGATTG    20400

ATGAAGCAGA TGCGATAGTG ATTGGAATTG GTGCAGGCAT GTCTGCATCT GACGGATTTA    20460

CATATGTAGG AGAGCGTTTT ACGGAAAATT TCCCAGATTT TATTGAAAAA TATCGCTTCT    20520

TTGATATGTT GCAAGCGAGT TTACATCCTT ATGGCAGTTG GCAAGAGTAT TGGGCATTTG    20580

AGAGTCGTTT TATTACATTA AACTATTTAG ATCAACCTGT AGGTCAGTCT TACCTCGCTT    20640

TAAAATCCTT GGTGGAAGGT AAACAGTACC ACATTATAAC TACGAATGCA GATAATGCTT    20700

TCGATGTAGC TGATTATGAT ATGACTCATG TATTTCATAT ACAAGGGGAG TATATACTGC    20760

AACAGTGTAG cTCAGCATTG TCATGCTCAA ACGTATCGCA ATGATGATTT AATTCGTAAA    20820

ATGGTTGTTG CGCAACAAGA TATGCTTATA CCTTGGGAGA TGATTCCAAG ATGTCCAAAA    20880

TGTGATGCCC CAATGGAAGT GAATAAACGT AAAGCGGAAG TTGGGATGGT TGAAGATGCT    20940

GAATTTCATG CGCAACTACA TCGTTATAAT GCTTTTCTAG AGCAACATCA AGATGATAAA    21000

GTGTTGTATT TGGAAATTGG AATTGGTTAT ACTACACCAC AATTTGTGAA GCATCCTTTT    21060

CAGCGTATGA CACGTAAAAA TGAAAATGCC CTTTATATGA CGATGAATAA AAAGGCATAT    21120

CGCATTCCGA ATTCAATTCA AGAACGTACC ATACATTTAA CTGAGGATAT CTCAACATTG    21180

ATTACAGCAG CACTCCGGAA CGACAGCACA ACGAAAAATA ACAACATTGG AGAGACAGAA    21240

GATGTACTTA ATAGAACCGA TTAGAAATGG AGAATATATT ACTGATGGTG CGATTGCACT    21300

CGCTATGCAA GTTTATGTTA ACCAGCATAT CTTTTTAGAT GAAGATATTT TATTCCCTTA    21360

TTATTGTGAT CCAAAAGTGG AAATTGGACG TTTTCAAAAT ACTGCTATAG AAGTGAATCA    21420

AGATTATATA GATAAACACA GTATTCAAGT AGTTCGCCGA GATACTGGTG GTGGCGCTGT    21480

GTATGTTGAT AAAGGTGCCG TTAATATGTG TTGTATTTTA GAACAAGACA CTTCAATTTA    21540

TGGTGATTTT CAACGATTTT ATCAACCAGC TATAAAGGCG TTGCATACAT TAGGTGCAAC    21600

AGATGTGGTA CAAAGCGGTA GAAATGATTT AACATTGAAT GGTAAAAAAG TGTCAGGCGC    21660

CGCAATGACA TTAATGAATA ATCGTATTTA TGGCGGTTAT TCGCTATTAC TTGATGTTAA    21720

TTATGAAGCA ATGGATAAAG TGTTAAAGCC TAATCGCAAA AAGATTGCAT CGAAAGGGAT    21780

TAAATCTGTG CGCGCACGTG TTGGTCATCT TAGAGAAGCA CTGGATGAAA AGTATCGTGA    21840

TATAACCATT GAAGAATTTA AAAATTTAAT GGTGACGCAG ATTTTGGGAA TCGATGACAT    21900

TAAAGAGGCG AAACGATATG AATTAACGGA TGCAGATTGG GAAGCGATTG ATGAATTAGC    21960

TGATAAAAAG TATAAAAATT GGGATTGGAA TTATGGCAAG TCACCCAAAT ATGAATACAA    22020

TCGAAGTGAA AGATTATCTT CAGGTACGGT AGACATAACA ATTTCTGTTG AACAAAATCG    22080
```

351

```
AGAAGCATTA CAAGGAACAA AAATGACAAG AGAAGATTTA ACGCATCAGT TAAAGCAATT    22200

AGACATCGTT TATTATTTTG GCAATGTTAC GGTAGAAGCA TTAGTGGATA TGATTTTAAG    22260

TTAATATTGT TATTTTATGT ATGCTGAATC ATTGGAAGTG TTTGCTTGCT CTTGAAAAGG    22320

TGACAATAGT GTTTGGTGAA GGTTGAACAT ATGAGTGGAA ATTATTGCCT TTAACTATTC    22380

AAAGTATGAT ATATATATGG TTTTTGTTTC TAAATGATTG GGTATTTGAA AATAGATGAG    22440

TTTAATATTT TAAGGAATAT AATGATGTTT ACTTTTATAA TTCATATAGA ATATTAAGCA    22500

ATATAAGTCT GTTGATATAT ACAAAATATA ATGACTGCTA TAATGAGTAA TCAATAGACA    22560

CAAAGAGGAG ATTATGTGAT GAATAATAAA GTATTAGTAA CCGGTGGTAC AGGGTTTGTT    22620

GGCATGCGAA TTATTTCACG ATTATTAGAA CAAGGTTATG ACGTACAAAC GACGATACGT    22680

GATTTAAGTA AAGCTGATAA AGTAATTAAA ACAATGCAAG ACAATGGCAT TTCCACAGAG    22740

CGATTAATGT TTGTCGAAGC GGATTTATCA CAAGATGAAC ATTGGGATGA AGCAATGAAA    22800

GATTGCAAGT ATGTCTTGAG TGTAGCATCT CCGGTGTTTT TCGGTAAAAC AGACGATGCA    22860

GAAGTGATGG CGAaCTGcAA TTGAAGGTAT ACAACGTATT TTAAGAGCTG CAGAACATGC    22920

GGGTGTTAAA CGTGTGGTAA TGACTGCAAA CTTTGGTGCA GTTGGTTTTA GTAATAAAGA    22980

TAAAAATTCA ATCACAAATG AAAGTCATTG GACAAATGAA GATGAACCAG GCTTATCAGT    23040

ATATGAAAAA TCAAAATTGT TAGCTGAAAA GGCAGCGTGG GATTTTGTTG AGAATGAAAA    23100

TACAACAGTA GAATTTGCCA CAATCAATCC AGTTGCAATT TTTGGGCCAT CATTAGATGC    23160

ACACGTTTCA GGAAGCTTTC ATTTATTAGA AAATTTATTG AATGGTTCAA TGAAACGTGT    23220

ACCGCAAATT CCGTTAAATG TTGTTGATGT GAGAGACGTA GCTGAACTGC ACATTTTGGC    23280

AATGACAAAT GAACAAGCTA ATGGCAAGCG ATTTATTGCG ACGGCTGATG GACmAATTwA    23340

tTTGTTGGGA ATTGcCAAAt TAATTAAAGA AAAGGGCCTG GAAATAGCTC CAAAAGTTCC    23400

TACTAAAAAA TTACCCAGCT TTATTTTGAG CnAnGnGCC                           23439
```

(2) INFORMATION FOR SEQ ID NO: 39:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4522 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

```
CCCTTTGAGA GTATATCATC TAGTCAAATT ATGCCTGTCA TTAGAGCGAC TAGCTTTGAT    60
```

```
TATTATGCAG TCGATTTAGG GAAATCATAT CGTCTAATTG ACGAAAGCAT GTTAGAGGAT    180

TTGAAGTTAA CTGAACAACA AATAAGAGAA ATGTCTCTGT TTAATGTTAG AAAATTGTCA    240

AATTCATATA CGACTGATGA AGTAAAAGGT AATATTTTTT ATTTTATTAA CTCAAATGAC    300

GGGTATGATG CAAGTAGGAT ACTAAATACT GCATTTTTAA ATGAAATTGA GGCACAATGT    360

CAAGGCGAAA TGCTCGTAGC AGTGCCACAC CAAGATGTGT TAATTATTGC AGATATACGC    420

AATAAAACAG GATATGATGT GATGGCACAT TTAACAATGG AATTTTTCAC TAAAGGTCTA    480

GTTCCAATTA CATCATTATC CTTTGGATAT AAACAGGGTC ATCTTGAACC GATATTTATT    540

TTAGGTAAAA ATAATAAACA AAAAAGAGAT CCAAACGTGA TTCAGCGTTT AGAAGCAAAT    600

CGTCGTAAAT TTAATAAAGA TAAATAGAAA TAATTGGATA AGGAGTTTTG TCATAATGAA    660

TTTATTTTAC AATCCTAAAT ATGTAGGAGA TGTCGCATTT TTACAAATTG AACCAGTTGA    720

AGGTGAATTA AACTACAATA AAAAAGGTAA TGTTGTTGAA ATTACtAATG AAGGTAATGT    780

TGTAGGTTAT AATATTTTTG AAATTTCAAA AGATATAACA ATTGAAGAAA AAGGTCATAT    840

TAAATTAACT GATGAACTTG TAAATGTATT CCAAAAGCGT ATTTCAGAAG CTGGTTTTGA    900

TTATAAATTA AATGCTGATC TATCACCGAA ATTTGTAGTT GGCTACGTTG AAACTAAAGA    960

CAAACATCCT GATGCAGATA AATTAAGTGT ACTAAATGTA AACGTTGGAA ATGACACATT   1020

ACAAATTGTA TGTGGCGCGC CTAACGTTGA AGCTGGACAG AAAGTTGTTG TTGCTAAAGT   1080

AGGTGCAGTG ATGCCTAGCG GTATGGTAAT TAAAGATGCT GAATTACGTG GTGTTGCCTC   1140

AAGCGGTATG ATTTGTTCAA TGAAAGAATT GAATTTACCT AATGCACCTG AAGAAAAAGG   1200

TATTATGGTA TTAAATGACA GCTATGAAAT TGGACAAGCA TTtTTTGAAT AATTAAGGAA   1260

GGTAGTGAAA ATATGAGCTG GTTTGATAAA TTATTCGGCG AAGATAATGA TTCAAATGAT   1320

GACTTGATTC ATAGAAAGAA AAAAAGACGT CAAGAATCAC AAAATATAGA TrACGATCAT   1380

GACTCATTAC TGCCTCAAAA TAATGATATT TATAGTCGTC CGAGGGGAAA ATTCCGTTTT   1440

CCTATGAGCG TAGCTTATGA AAATGAAAAT GTTGAACAAT CTGCAGATAC TATTTCAGAT   1500

GAAAAAGAAC AATACCATCG AGACTATCGC AAACAAAGCC ACGATTCTCG TTCACAAAAA   1560

CGACATCGCC GTAGAAGAAA TCAAACAACT GAAGAACAAA ATTATAGTGA ACAACGTGGG   1620

AATTCTAAAA TATCACAGCA AAGTATAAAA TATAAAGATC ATTCACATTA CCATACGAAT   1680

AAGCCAGGTA CATATGTTTC TGCAATTAAT GGTATTGAGA AGGAAACGCA CAAGCCAAAA   1740

ACACATAATA TGTATTCTAA TAATACAAAT CATCGTGCTA AAGATTCAAC TCCAGATTAT   1800

CACAAAGAAA GTTTCAAGAC TTCAGAGGTA CCGTCAGCTA TTTTTGGCAC AATGAAACCT   1860
```

```
AAACAAAAAT ATGATAAATA TGTAGCTAAG ACGCAAACGT CTCAAAATAA ACAATTAGAA          1980

CAAGAAAAAC AAAATGATAG TGTTGTCAAA CAAGGAACTG CATCTAAATC ATCTGATGAA          2040

AATGTATCAT CAACAACAAA ATCAATGCCT AATTATTCAA AAGTTGATAA TACTATCAAA          2100

ATTGAAAATA TTTATGCTTC ACAAATTGTT GAAGAAATTA GACGTGAACG AGAACGTAAA          2160

GTGCTTCAAA AGCGTCGATT TAAAAAAGCG TTGCAACAAA AGCGTGAAGA ACATAAAAAC          2220

GAAGAGCAAG ATGCAATACA ACGTGCAATT GATGAAATGT ATGCTAAACA AGcGGAACgC          2280

TATGTTGGTG ATAGTTCATT AAATGATGAT AGTGACTTAA CAGATAATAG TACAGATGCT          2340

AGTCAGCTTC ATACAAATGG CATAGAGAAT GAAACTGTAT CAAATGATGA AAATAAACAA          2400

GCGTCAATAC AAAATGAAGA CACTAATGAC ACTCATGTAG ATGAAAGTCC ATACAATTAT          2460

GAGGAAGTTA GTTTGAaTCA AGTATCGACA ACAAAACAAT TGTCAGATGA TGAAGTTACG          2520

GTTTCGAATG TAACGTCTCA ACATCAATCA GCACTACAAC ATAACGTTGA AGTAAATGAT          2580

AAAGATGAAC TAAAAAATCA ATCCAGATTA ATTGCTGATT CAGAAGAAGA TGGAGCAACG          2640

aATAAAGAAG AATATTCAGk AAGTCAAATC GATGATGCAG AATTTTATGA ATTAAATGAT          2700

ACAGAAGTAG ATGAGGATAC TACTTCAAAT ATCGAAGATA ATACCAATAG AAACGCGTCT          2760

GAAATGCATG TAGACGCTCC TAAAACGCAA GAGTACGCAG TAACTGAATC TCAAGTAAAT          2820

AATATCGATA AAACGGTTGA TAATGAAATT GAATTAGCAC CGCGTCATAA AAAAGATGAC          2880

CAAACAAACT TAAGTGTCAA CTCATTGAAA ACGAATGATG TGAATGATAA TCATGTTGTG          2940

GAAGATTCAA GCATGAATGA AATAGAAAAG AATAACGCAG AAATTACAGA AAATGTGCAA          3000

AACGAAGCAG CTGAAAGTGA ACAAAATGTC GAAGAGAAAA CTATTGAAAA CGTAAATCCA          3060

AAGAAACAGA CTGAAAAGGT TTCAACTTTA AGTAAAAGAC CATTTAATGT TGTCATGACG          3120

CCATCTGATA AAAAGCGTAT GATGGATCGT AAAAAGCATT CAAAAGTCAA TGTGCCTGAA          3180

TTAAAGCCTG TACAAAGTAA GCAAGCTGTG AGTGAAAGAA TGCCTGCGAG TCAAGCCACA          3240

CCATCATCAA GATCTGATTC ACAAGAGTCA AATACAAATG CATATAAAAC AAATAATATG          3300

ACATCAAACA ATGTTGaGAA CAATCAACTT ATTGGTCATG CAGAAACAGA AAATGATTAT          3360

CAAAATGCAC AACAATATTC AGAGCAGAAA CCTTCTGTTG aTTCAACTCA AACGGAAATA          3420

TTTGAAGAAA GTCAAGATGA TAATCAATTG GAAAATGAGC AAGTTGATCA ATCAACTTCG          3480

TCTTCAGTTT CAGAAGTAAG CGACATAACT GAAGAAAGCG AAGAACAAC ACATCCAAAC          3540

AATACTAGTG ACAACAAGA TAATGATGAT CAACAAAAAG ATTTACAGTC ATCATTTTCA          3600

AATAAAAATG AAGATACAGC TAATGAAAAT AGACCTCGGA CGAACCAACA AGATGTTGCA          3660
```

354

CCAAGTGTTT CATTACTAGA AGAACCACAA GTTATTGAGT CGGACGAGGA CTGGATTACA   3780

GATAAAAAGA AAGAACTGAA TGACGCATTA TTTTACTTTA ATGTACCTGC AGAAGTACAA   3840

GATGTAACTG AAGGTCCAAG TGTTACAAGA TTTGAATTAT CAGTTGAAAA AGGTGTTAAA   3900

GTTTCAAGAA TTACGGCATT ACAAGATGAC ATTAAAATGG CATTGGCAGC GAAAGATATT   3960

CGTATAGAAG CGCCTATTCC AGGAACTAGT CGTGTTGGTA TTGAAGTTCC GAACCAAAAT   4020

CCAACGACAG TCAACTTACG TTCTATTATT GAATCTCCaA GTTTTAAAAA TGCTGAATCT   4080

AAATTAACAG TTGCGATGGG GTATAGAATT AATAATGAAC CATTACTTAT GGATATTGCT   4140

AAAACGCCAC ACGCACTAAT TGCAGGTGCA ACTGGATCAG GGAAATCAGT TTGTATCAAT   4200

AGTATTTTGA TGTCTTTACT ATATAAAAAT CATCCTGAGG AATTAAGATT ATTACTTATC   4260

GATCCAAAAA TGGTTGAATT AGCTCCTTAT AATGGTTTGC CACATTTAGT TGCACCGGTA   4320

ATTACAGATG TCAAAGCAGC TACACAGAGT TTAAAATGGG CCGTAGAAGA AATGGAACGA   4380

CGTTATAAGT TATTTGCACA TTACCCATGT ACGTAnTATA ACAGCATTTA ACnAAAAGC   4440

CCCATATGAT GAAAGAATGn CAAAAATTGT CATTGTAaTT GATGAGTTGG CTGATTTAAT   4500

GATGATGGTC CGCAAGAAGT TG   4522

(2) INFORMATION FOR SEQ ID NO: 40:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 751 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TCAAGTTTAC GGATACGTAT ATATTTTGCA TGACATTTAG TGCAATAATA TTCATAATTT   60

GCCCGTTGTT GATAGCTTTC AATGCTGTTA CAAAATCTAG GCGCTCCAAC CTGTTGGCTC   120

AATCGTTTAA AATCTTGATC TTTATGTTGA TAACCTTTAC CAGCAATATG CAAGTGATAA   180

TGACACAATT CGTGCAGTAT AATTTTTACA ACAGCATCTT CTCCATAATG CTCATATTGT   240

TTTGGATTAA TTTCAATATC ATGGGACTTT AAAAGATAAC GTCCGCCTGT TGTACGTAAC   300

CTTTTATTAA AATATGCACA ATGTCGAAAC GTACGTCCAA ATTTTTCTTC CGAAAGATTC   360

TCAACCATTC GCTGAAGTTT GTCATTATTC ATGTGGATCA ATCATCGTTA ATGATACTTT   420

GTCTTTATTT TTGTCAATAC TGTAAATCCA AACGTCAACG ATATCACCAA CACTGACAAT   480

ATCCATTGGA TTTTTTACGA ACTTCTTAGA AAGTTTCGAA ACATGGACAA GTCCATCTTG   540

TTTCATTCCT TCTTGTAAAT CTTCAATTGA TAGCACATCG GATTTAAGGA TTGGTGTTTC   660

AAACTCGTCC CTTGGATCTC GATTAGGTGC GTTCAAGGAT TTAATAATAT CCTCTAATGT   720

AGGTACACCG ACTTGTAATT CAATCGCCAG T   751

(2) INFORMATION FOR SEQ ID NO: 41:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1076 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TCTCCAGCTT TAACTTGATC TGGCACTTTA ACAATTGTCT GATCCATACA TACGCGACCA   60

ATAACTTCGC ATTGATGACC ATTTACATTT ACAAAGCTAC CTTGCATTAT GCGTAAATGG   120

CCATCTGCAT ATCCAATAgG TAACAATGCT ATTGTAGTTG GGTCAGTAGC TGTATAAGTT   180

GCACCATAAC TTACAGACTC ACCCGCTTGT AGCGTCTTTG TTTGAACTAC ATTAGCAATT   240

AATTGCACAC TTGGTTTAAG GTGTACTTTA ACTTTTTGCT GTACATACTC TGATGGATAA   300

TATCCATAAA GGGAAATTCC TGGTCTTATT GCATTACAGA ATTGGCAATC CATTAATAGA   360

GAGCCTGCTG AGTTCTGACA ATGTATATAT TCAGGTTTAA TTGCTTCATT GACCATATCT   420

TTAAAACGTT GATATTGTTC AGTTGTCATA TCTCCTGGTT CGTCAGCACA GGCAAAGTGT   480

GTAAACACGC CTTCAAATAC AAGTTGCTCA TATTGTTGAA TGATTTCAAT CACTTCTTGA   540

TACGTTTTAG TATCTTTAAT ACCTAAACGT CCCATTCCTG TATCTAATTT AATGTGCAAC   600

CATAACTTTT TCTCTTGCTC ACCAGAAATG TTTTTAATTG CTTCTTTCAA CCACTGTTTA   660

GACGGAACCG TTAAGGCAAC TCGGTGTTGT ATCGCTTTAT CAATATCTTT AGCTGGTAAC   720

ACACCTAAGA CTAAAATTTT AGCAGTAATC CCATGCATTC TAAGTTCTAT CGCTTCATCT   780

AACGTTGCTA CAGCAAAAAA TGTGGCGCCA TTTTCCATTA AATGACGTGC TACTTTAACA   840

CTACCTAGTC CATAGGCATT GGCTTTAACG ACAGCCATCA CTGTTTTATT TGGATGCAAT   900

GTACTGAATA CTTTGAAATT TGATGCAACA GCGTTTAAAT CTACATTCAT ATACGCAGAT   960

CTATAATATT TATCCGACAT ATTACTTCCT CCTGTAATTC CCACACGTTT TAAAACTAGA   1020

TCTTAATTAT CATTGTATAA CAAATTTAAA ATGCTGACTT TTCTAAAACA ACTTGG   1076

(2) INFORMATION FOR SEQ ID NO: 42:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 2930 base pairs

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

```
TGACCACAAT GCCCAATACA ACCATCCCAT GGTAAAGCCA AGAGATGAGT CAATAAAGCG      60

TGTTGAATAA GAGCTGAATG AACCTGATAC TGGATAAAAT GTTGCCAACT CTCCAATTGA     120

TGACATTAAG AAATATAGCA TGACACCAAT AACAAGATAA GCGAGTATAG CGCCTCCAGG     180

ACCAGCTTGA GAAATGATAT TACCAGTAGC TACAAATAGA CCAGTCCCAA TTGCACCACC     240

TATAGCAATC ATGGAAATGT GTCTTGAGTT AAGACTACGG TTCATTTTAT TATCTTCCAT     300

ATTTAGTCTC CCATCTATTT AAATATACCC ATTATTGTAA GCTTTTTAAG TGTACTATTC     360

AATAACTATT TAGTACTGTA AAGCGAAAAA ATTAAAATTT TCTGATTTTT TAATCATCTT     420

GAGCATGTTT AATTGTAATT TTGATGGGGT TAAATTATAA TATGTATTAA ATTATAATTA     480

TnATAAATTG TGGAGGGaTG ACTATGTCAC AACAAGACAA AAAGTTAACT GGTGTTTTTG     540

GGCATCCAGT ATCAGACCGA GAAAATAGTA TGACAGCAGG GCCTAGGGGA CCTCTTTTAA     600

TGCAAGATAT TTACTTTTTA GAGCAAATGT CTCAATTTGA TAGAGAAGTA ATACCAGAAC     660

GTCGAATGCA TGCCAAAGGT TCTGGTGCAT TTGGGACATT TACTGTAACT AAAGATATAA     720

CAAAATATAC GAATGCTAAA AtATTCTCTG AAATAGGTAA GCAAACCGAA ATGTTTGCCC     780

GTTTCTCTAC TGTAGCAGGA GAACGTGGTG CTGCTGATGC GGAcGTGACA TTCGAGGATT     840

TGCGTTAAAG TTCTACACTG AAGAAGGGAA CTGGGaTTTA GTAGGGAATA ACACACCaGT     900

ATTCTTCTTT AGAGATCCAA AGTTATTTGT TAGTTTAAAT CGTGCGGTGA AACGAGATCC     960

TAGAACAAAT ATGAGAGATG CACAAAATAA CTGGGATTTC TGGaCGGGTt TCCAGAAGCA    1020

TTGCACCAAG TAACGATCTT AATGTCAGAT AGAGGGATTC CTAAAGATTT ACGTCATATG    1080

CATGGGTTCG GTTCTCACAC ATACTCTATG TATAATGATT CTGGTGAACG TGTTTGGGTT    1140

AAATTCCATT TTAGAACGCA ACAAGGTATT GAAAACTTAA CTGATGAAGA AGCTGCTGAA    1200

ATTATAGCTA CAGATCGTGA TTCATCTCAA CGCGATTTAT TCGAAGCCAT TGAAAAAGGT    1260

GATTATCCAA AATGGACAAT GTATATTCAA GTAATGACTG AGGAACAAGC TAAAAACCAT    1320

AAAGATAATC CATTTGATTT AACAAAAGTA TGGTATCACG ATGAGTATCC TCTAATTGAA    1380

GTTGGAGAGT TTGAATTAAA TAGAAATCCA GATAATTACT TTATGGATGT TGAACAAGCT    1440

GCGTTTGCAC CAACTAATAT TATTCCAGGA TTAGATTTTT CTCCAGACAA AATGCTGCAA    1500

GGGCGTTTAT TCTCATATGG CGATGCGCAA AGATATCGAT TAGGAGTTAA TCATTGGCAG    1560
```

GGTCAAATGC GCGTAGTTGA CAATAACCAA GGTGGAGGAA CACATTATTA TCCAAATAAC   1680

CATGGTAAAT TTGATTCTCA ACCTGAATAT AAAAAGCCAC CATTCCCAAC TGATGGATAC   1740

GGCTATGAAT ATAATCAACG TCAAGATGAT GATAATTATT TTGAACAACC AGGTAAATTG   1800

TTTAGATTAC AATCAGAGGA CGCTAAAGAA AGAATTTTTA CAAATACAGC AAATGCAATG   1860

GAAGGCGTAA CGGATGATGT TAAACGACGT CATATTCGTC ATTGTTACAA AGCTGACCCA   1920

GAATATGGTA AAGGTGTTGC AAAAGCATTA GGTATTGATA TAAATTCTAT TGATCTTGAA   1980

ACTGAAAATG ATGAAACATA CGAAAACTTT GAAAAATAAA TTTGATATGT AGTTTCTATA   2040

TTGCGTAGTT GAGCAGTTTA TGATATCATA ATAAATCGTA AAGATTCCTA ACAAGAGAGG   2100

GTGTTTAACG TGCGCGTAAA CGTAACATTA GCATGCACAG AATGTGGCGA TCGTAACTAT   2160

ATCACTACTA AAAATAAACG TAATAATCCT GAGCGTATTG AAATGAAAAA ATATTGCCCA   2220

AGATTAAACA AATATACGTT ACATCGTGAA ACTAAGTAAT TCTTATCATT CAAATACGAC   2280

GATTTGAAAA TAAAGCGGGC TTACCTATTA TATTGGGGAG CTCGCTTTTT TATGAAATTT   2340

TTGTGAAGAG TGATTAATGG ATTGAGTTTC ATCGGTAGAA CAATATATGA TTATATTAGT   2400

TGTTACTTTA TTAAAaTTTG AGAATATTTA TAGAAGGAAA TAGATTACTG ATTTTATAAA   2460

GTCACTTTGT TAGCGAATGC TTGAAAGAGT ATTTAATATA GTAGAATTTA AAATTTCAAA   2520

GCGGAATTTA ATAAGTACGA AGTAGTTCTG GGTATGTTTT ATAAATGTTC GATAATACAC   2580

TTTAATCTTA AATATGATGG TTTAGAAAAT GATTTAACAA AGAAATGAaA CTTTACTGTT   2640

GAATTATGTG AGGATTGTGT TATTATATAA ATCGTAATAA TTACGATTTG ATAAAAAGTG   2700

AGGTAACTAT ATATGGCTAA GAAATCTAAA ATAGCAAAAG AGAGAAAAAG AGAAGAGTTA   2760

GTAAATAAAT ATTACGAATT ACGTAAAGAG TTAAAAGCAA AAGGTGATTA CGAAGCGTTA   2820

AGAAAATTAC CAAGAGATTC ATCACCTACA CGTTTAACTA GAAGATGTAA AGTAACTGGA   2880

AGACCTAGAG GTGTATTACG TAAATTTGAA ATGTCTCGTA TTGCGTTTAG   2930

(2) INFORMATION FOR SEQ ID NO: 43:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3606 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

CTTCTTGCCA TGGCTCTCTT TATTTAAAAA TGCTTCCAAC TTGTCCATTT GATTGTTTCT   60

```
TTATAAAAAA  CTAATTTTAC  AAATGCTTTT  GCGTTCTTAC  AAAAAATGCA  TTTGACTATT      180

ATTATAATAA  GCGTATAATT  GTCGCATATT  ATTTTTTGTA  TTTTTGGCAA  TAACGAAGGA      240

GTATTTATGA  ATAAAGACAA  GCAATTGCAC  AACGACAAAA  TCAATCTATC  CCAATTAGTC      300

TTATTAGGGT  TAGGCTCTTT  AATAGGATCT  GGTTGGCTAT  TTGGTGCGTG  GGAAGCATCA      360

TCAATAGCTG  GACCAGCAGC  AATCATATCA  TGGGTTCTTG  GATTCCTAGT  CATTGGAACC      420

ATTGCCTATA  ACTACATTGA  AATCGGCACA  ATGTTTCCTC  AATCAGGTGG  CATGAGTAAC      480

TATGCCCAGT  ATACACATGG  CTCATTATTA  GGCTTTATTG  CTGCTTGGGC  GAATTGGGTG      540

TCTTTGGTGA  CAATAATACC  TATCGAAGCT  GTGTCAGCTG  TTCAATATAT  GAGTTCTTGG      600

CCGTGGCATT  GGGCGAAACC  AATGAGATAT  TTAATGGAAA  ATGGCTCTAT  TAGCACATAC      660

GGATTGCTAG  CTGTATATCT  CATCATTGTT  ATTTTTTCAT  TATTAAACTA  TTGGTCCGTA      720

AAACTTTTAA  CATCATTTAC  GAGTTTAATT  TCTGTATTTA  AATTAGGCGT  ACCCATGTTA      780

ACCATCATCA  TGTTGATGCT  ATCAGGATTC  GACACTTCAA  ATTACGGCCA  TTCGGCAAGC      840

ACATTTATGC  CTTACGGAAG  TGCACCGATT  TTTGCTGCAA  CAACAGCATC  AGGGATTATT      900

TTTTCATTCA  ATTCATTCCA  GACAATTATT  AATATGGGTT  CAGAAATTAA  AAATCCTGAA      960

AAAAATATCG  CAAGAGGCAT  CGCTATCTCA  CTGTCAATCA  GTGCAGTGTT  GTACATCATT     1020

TTACAAAGTA  CGTTTATCAC  TTCTATGCCT  CAATCAATGT  TACAACATAG  TGGATGGAAT     1080

GGCATCAACT  TCAATTCACC  ATTTGCTGAT  TTAGCTATCT  TATTAGGAAT  TAATTGGCTC     1140

GCAATTTTAC  TATACATTGA  AGCTTTTGTA  TCACCATTCG  GTACTGGCGT  GTCATTTGTC     1200

GCCGTTACAG  GTCGAGTTTT  ACGAGCAATG  GAGAAAAATG  GACATATCCC  TAAATTTCTT     1260

GGGAAGATGA  ATGAAAAATA  TCATATCCCA  CGTGTAGCAA  TCATCTTTAA  TGCCATCATT     1320

AGTATGATTA  TGGTTACATT  ATTTAGAGAT  TGGGGTACGC  TAGCAGCAGT  TATTTCTACT     1380

GCAACTTTAG  TAGCCTATTT  AACTGGCCCA  ACGACAGTGA  TTGCATTAAG  AAAAATGGGA     1440

CCAACAATGA  CTCGTCCATT  TAGAGCAAAA  ATTTTAAAAG  TAATGGCACC  ATTATCATTT     1500

GTATTAGCTT  CATTAGCTAT  ATATTGGGCA  ATGTGGCCAA  CAACGGCTGA  AGTTATTTTA     1560

ATCATTATAC  TTGGATTACC  AATCTACTTC  TTCTATGAAT  ATCGTATGAA  TTGGCGTAAT     1620

ACAAAGAAAC  AAATTGGTGG  TAGCTTATGG  ATTATTGTAT  ATTTAATCGT  GCTATCAATA     1680

CTGTCATTTA  TAGGAAGCAA  AGAATTTAAA  GGCTTAAATA  TGATTCACTA  TCCATTTGAC     1740

TTTATCGTTA  TTATTATTGT  GGCACTTATC  TTCTATTACA  TCGGTACAAC  GAGTTCATTT     1800

GAAAGCGTCT  ATTTCCGTCG  CGCAACACGA  ATCAATACGA  AGATGCGTGA  GTCACTAAAT     1860
```

CACACACATT AACCAACCAT TGATTTCAAC ATCTTGGTTG GTTTTTTATT TTGAAAATCG   1980

GTTATAAATA ACTAACATAA CAAGATGATG ATCAGGCTGG GACATAAATC AATGTTCTAT   2040

GCTCTACGAA gTTATATTGG CAGTAGTTGA CTGAACGAAA ATGCGCTTGT AACAAGCTTT   2100

TTTCGATTCT AGTCAGGGGC CCCAACACAG AGAATTTCGA AAAGAAATTC TACAGGCAAT   2160

GCAAGTTGGG GTGGGACGAC GATAAAGAAA TACTTTTTCT ATAGAAATTA GTATYtCTTA   2220

TGCATGAGTT TTACTCATGT ATTCATATTT TTAAGTACAC ATTAGCTGTG GCTAATGTAT   2280

AAGAACCACT ACATAATAAA TCATTTGTGG CTCTTTATCA TTTCTGTCCC ACTCCCGTAG   2340

AAGTACATCA TATAATGCTG AAAATGGTTT GAGTTAAAAC AGATATCAAG CTCGTCTGAT   2400

TCAGTCACAA AATTGTCTTG TTATACTTGT CACCTATCAT CTATAGACCG TGGTATGATT   2460

AAATTGGGGA TGATAAAGGA GGTTAATAAA TATGAAGATT AATACTACAG GTGGTCAAAT   2520

TCATGGTATT ACACAAGATG GTTTAGATAT CTTCTTAGGC ATTCCTTATG CAGAACCACC   2580

AGTTCATGAC AATCGCTTTA AACATTCTAC GTTAAAAACA CAATGGTCAG AGCCAATTGA   2640

TGCAACTGAA ATACAACCCA TCCCACCGCA ACCAGACAAC AAATTAGAAG ATTTTTTCTC   2700

CTCACAATCT ACAACTTTTA CTGAACATGA AGACTGTTTA TATCTAAATA TTTGGAAACA   2760

ACATAATGAT CAGACGAAGA AACCTGTCAT CATTTATTTT TATGGTGGTA GTTTTGAAAA   2820

TGGTCATGGT ACAGCCGAAC TCTATCAACC GGCACATTTA GTACAAAATA ACGACATTAT   2880

CGTTATTACA TGCAATTATC GTTTAGGCGC ATTAGGATAT TTAGACTGGT CATATTTTAA   2940

TAAAGATTTT CATTCCAATA ATGGCCTTTC AGATCAAATC AATGTCATAA AATGGGTGCA   3000

TCAATTTATT GAATCCTTCG GTGGCGACGC TAATAACATT ACTTTAATGG GTCAGTCTGC   3060

AGGCAGTATG AGCATTTTGA CTTTACTTAA AATACCTGAC ATTGAGCCAT ACTTCCATAA   3120

AGTCGTTCTA CTAAGTGGCG CACTACGATT AGACACCCTT GAGAGTGCAC GCAATAAAGC   3180

ACAACATTTC CAAAAAATGA TGCTCGATTA TTTAGATACA GATGATGTTA CATCATTATC   3240

GACAAATGAT ATTCTTATGC TGATGGCGAA gcTAAAACAA TCTCGAGGAC CTTCTAAAGG   3300

GCTTGATTTA ATATATGCGC CTATTAAAAC AGATTATATA CAAAATAATT ATCCAACAAC   3360

GAAACCAATT TTTGCATGTT ATACAAAAGA TGAAGGCGAT ATTTATATTA CTAGTGAACA   3420

GAAAAAATTA TCGCCGCAAC GCTTTATCGA CATTATGGAA TTAAATGATA TTCCTTTAAA   3480

ATACGAAGAT GTTCAGACGG CGAAGcAACA ATCTTTAGCG ATTACACATT GTTATTTCaA   3540

ACAGCCGATG aAGCAATTTT TACmACmACT CAATATACmA GATTCCAACC GCACCAACTA   3600

TGGCTT   3606

```
          (i)  SEQUENCE CHARACTERISTICS:
               (A)  LENGTH: 15109 base pairs
               (B)  TYPE: nucleic acid
               (C)  STRANDEDNESS: double
  5            (D)  TOPOLOGY: linear
```

```
          (xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 44:

 10   GAAATTAAAA AAGCAATTGG nACAAGATGC AACAGTGTCA TTGTTTGATG AATTTGATAA         60

      AAAATTATAC ACTTACGGCG ATAACTGGGG TCGTGGTGGA GAAGTATTAT ATCAAGCATT        120

      TGGTTTGAAA ATGCAACsAG AACAACAAAA GTTAACTGCA AAAGCAGGTT GGGCTGAAGT        180
 15
      GAAACAAGAA GAAATTGAAA AATATGCTGG TGATTACATT GTGAGTACAA GTGAAGGTAA        240

      ACCTACACCA GGATACGAAT CAACAAACAT GTGGaAGAAT TTGAAAGCTA CTAAAGAAGG        300

      ACATATTGTT AAAGTTGATG CTGGTACATA CTGGTACAAC GATCCTTATA CATTAGATTT        360
 20
      CATGCGTAAA GATTTAAAAG AmAAATTAAT TAAAGCTGCA AAATAATTCA GCTATATAAG        420

      TTAGTGAAAT GAGAGTCTGA AACATATCAA TCTTTTGATA TTGTATTAGG CTCTTATTTT        480

      TATAGCTAGA AAGTTAGATA TTTGTATTTT TTTAAATAAT AAGTGCCGTT GTTATCGTTC        540
 25
      AATTTAATTA ATGATAGATT AGTATTATTA TAGCTAAAGT AGTATACCTG AGAAAATAGC        600

      TCAATGTATC TCTTTATTAA TAAGTTATAT CATAATTATT TTAGTGCATA CTTTATGGAA        660

      GGGATATCAG GGAATGGCTT TCAATTAAAG AAGAGGTTTA AAAGGATTAC AACAGAATGT        720
 30
      TATGATTTTG TAGAAAGATA TATAACAACG TTTTATAAAA ACATAATATT GTTAATGGAA        780

      AATGAAATGT AAGGGGGATT TCGAGTGACT AAGAAAGTTT ATTTTAACCA CGATGGTGGT        840

      GTAGATGATT TAGTATCTCT ATTTTTATTA TTACAAATGG AAAACGTTCA ATTGATAGGG        900
 35
      GTCAGTACAA TTGGTGCTGA TTGTTATTTA GAGCCATCTT TGAGCGCATC AGTAAAAATT        960

      ATTAATCGTT TTTCAAATGA AGATATTCAA GTTGCGCCAT CATATGAACG AGGAAAAAAT       1020

      CCATTTCCTA AAGAATGGCG TATGCATGCC TTTTTTATGG ACGCATTGCC AATTTTAAAT       1080
 40
      GAGCCAGTCA AACATGTTGC TTCAAATGTG AGCGACAAAG AAGCCTTTGA AGACATTATT       1140

      CAAACTTTAA AGAGACAATC AGAAAAAGTA ACATTATTAT TTACAGGCCC GCTTACAGAT       1200

      TTAGCAAAAG CACTACAAAA AGATTCATCT ATCGTTCAGT ATATAGAAAA ATTAGTTTGG       1260
 45
      ATGGGTGGCA CCTTTTTACC AAAAGGAAAT GTTGAAGAAC CTGAGCATGA TGGTTCTGCA       1320

      GAATGGAATG CATATTGGGA TCCAGAAGCG GTTAAAATTG TTTTTGATAG CGATATAGAG       1380
 50
      ATTGATATGG TTGCTTTAGA AAGTACGAAT CAAGTACCGC TAACGTTAGA TGTTAGACAA       1440
```

```
 55
```

```
GTACCACCAT TAACACACTT TATAACAAAT TCTACTTACT TTTTATGGGA TGTTTTAACG        1560

ACTGCTTATA TTGGTAACAA GGACTTGGTT CATTCAATTG AGAAAAAAGT CGATGTAATA        1620

AGTTATGGAC CAAGTCAAGG TAAGACATTT GAGTGTAAAG ATGGGCGCAA AATTAATGTC        1680

ATAAATCATG TAGATAACAA CGCATTTTTT GATTATATAA CTGCACTTGC TAAAAAAGTA        1740

AATTAACAGC TGTGTAGAAT AATTAAGGTT TTAATTTATA TAGAACAACT TATTGTAAAC        1800

TTTTCATTTC TTAAAGTTTA CAATGGTGCT ATAATAATGG TCATGAAATA CGAAAGGAAG        1860

TAAAAAATGA CAACAAAACA GTTAGTATAT ACAGCTTTAA TGACAGCGAT TATCGCTATT        1920

TTAGGATTGG TACCGGTAAT TCCACTACCA TTTTCTTCAG TACCAATTGT ACTTCAAAAC        1980

ATTGGTATTT TCTTAGCAGG TGCGATTTTA GGACGTAAAT ATGGCACATT AAGTGTTATC        2040

GTCTTTTTAT TATTAGTAGT TGCTGGCTTG CCATTGTTAT CAGGTGGTCG CGGTGGCATC        2100

GGTGTATTCG CAGGTCCTTC AGCAGGGTTT TTACTATTAT ATCCAGTTGT AGCATTCATG        2160

ATTGGGGCGA TTCGAGATAG ATTCATCAAT GAAATTAATT TCTGGATTTT ATTCGTTGGT        2220

ATTTTAGTTT TTGGTGTTAT AGCATTAGAT GTTATTGGTA CATTGATTAT GGGCATGATT        2280

ATTAACATAC CATTTACGAA AGCTATTTCA ATTTCATTAG CTTATTTGCC TGGTGATATA        2340

TTAAAAGCAA TTGTAGCAAG TTTGATTGGT ACAGCTTTAC TTAATCACTC GCAGTTTCGT        2400

CAAATTATGG GAATAAAATA ATCATATTTA AGATAGTAAA GTAATTGAAT AAGTTGCTTT        2460

GAAATTTATA AAAGTGAAAG GAGTAGGTGT CAATGGCTAG TATAAGTATG TCAGATATAT        2520

ATTGTAACGG CACTATATTT GAAAATGACG ACGAGCAGTT GATTTATTTA ACGCCTTCTT        2580

TTCCACAACG ATACACAAGT AACACATGGA TATATAAAAA GACGCCTACC CAAGAGCGAT        2640

GGCTGAAAGA CTTAGAACGT CAACATCAAT TACATACAAA TCAAGGTTCA AATCATTATG        2700

CGTTTAGTTT CCCGGAAAAT GAACAACTTG ATAATCATTG GATGGCTATG TTTAAAGATA        2760

TGAATTTTGA ACTAGGTATT ATGGAATTGT ATGCCATAGA AAGTGATGCG CTTGCCAATT        2820

TGCCGCGTAA CTCTGACGTT GAAATTGCCA TCGTTGACGA GTCGCATATA GATGCCTATT        2880

TAAAAGTTGC ATATCAGTTT AGTTTGCCAT TTGGAAAAGA CTATGCAGAT GCACATGAAG        2940

AAATGGTAAG GGAACATTAT CAAAAAGATG TGATTAAACG CTTAGTAGCT TATTTAAATA        3000

ATGAACCTAT TGGCGTTGTA GATGTCATTG AAAGTGAAAA TTACATTGAA TTAGATGGAT        3060

TTGGTGTATT AGAACAATTT CGGCACCAAG GAATTGGATC TACAATTCAA TCGTTGATAG        3120

GTGAATACGC CATATCAAAA AATCACAAAC CAATCATATT AGTTGCAGAT GGTGAAGATA        3180

CAGCAAAAGA TATGTATGCA AAGCAAGGTT ATGTCTATCA ATCGTTTTGT TATCAAATAT        3240
```

```
TAAGCTGGTT TCGAGTAGAA ATCAACTTAC TGCTTTTTAA ATTGTTTTGA GCTACTTATA    3360

CTTATAAAAA TAGTGCGTTT AAATTGTTGA TTCATGTAGA ATATCGTTCA TTATGACACA    3420

CTATAATGAA TATGTTATTG TTCAGAATCA ATGATACGTT CTGGATGACT GTATATATTA    3480

AAGCCACCAT TTCGAATAAA TCCAACTGCC GTAATATTTA GGTCATTAGC TAAGGTTACA    3540

GCAAGCGTTG TCGGAGCTGA TTTAGATAAA ATGACGCCAA CACCAATTTT TGCGGCTTTA    3600

ATTAAAATTT CTGATGAAAT ACGTCCACTA AAAATTAATA CTTTATCTCG GACAGTAATA    3660

TGTCGCTGAA TACAAAATCC ATATAATTTA TCTAGAGCGT TATGTCTACC AATGTCTTGT    3720

CGATGTACAA AAAATGTCAA ACCATCGCTT ATAGCAGCAT TATGTAAGCC ACCTGTTTCT    3780

TGGTAAATAT GACTTGCACT TTGTAATCGA GTCATCATGT TAATAATTTG CATTGGAGTT    3840

AAAGTGATTT TAGACATAGA TGTTTTAGCG ATAGCAGCAT CATTTTGAAA ATAAAACTCA    3900

CGACTCTTTC CGCAACAAGA TGCAATCATT CGTTTGTGG AATATTGAAA GCGATCGCCT    3960

AAATCTTTAT TAAGTTCAAC ATGGGCAAAA CCTTTACTAT CATCAATCAG TACAGATTTT    4020

AATTCATCTC GCTTTAAAAT GGCACCTTCC GAAGCCAGAA ATCCAATGAC TAACTCCTCA    4080

AGGTTTGTTG GACTGCATAT AACAGTCGCA AATTCTTCAC CATTCACCAT AATTGTAAGT    4140

GGAAATTCTG TCACATATTG ATCTGTTGTA TTGAATAATT TTCCATCTTC ATATCTAACA    4200

ATTGGTTGAC CTAAAGATAC ATCTTTGTTC ATTATCTAAC CCCTTTAATT AGCTTAAACT    4260

TTATTTTAAA GCAATTTGCT TAAAATTTTA ACATATTTGC TTAAGTTTGA AATTTGATTG    4320

ATAAAAATTA ATAGCGAGCA ATCTGTTTGA TTTAAATTGA ATTCGAGAAT ATACATACTA    4380

GGGCATCAAT TAATAAATAT CAATCTTATG CAAATTTGAC AATTGTTTGA ATCAATATAT    4440

AAACAGGCAA CGGTTCTTTT CAAATATAAT AGTAAGTGTA TAATGAAAAT GTAAATATTA    4500

TTAAAAATGG GGGTTCACTC AATGAAATTG AAACGTTTAT TTGCTGTTGT GATTGCAATG    4560

CTTTTAGTAT TAGCTGGTTG CTCTAATTCT AACGATAATA ATGAAAGTAA AAAAGATGAC    4620

GCAGACAATG GTAAGAAACA AGAGATTCAA GTTGCAGCGG CAGCAAGTTT AACAGATGTA    4680

ACCAAGAAAT TAGCTTCAGA ATTTAAAAAA GAGCATAAAA ATGCTGATAT TAAATTTAAC    4740

TATGGTGGAT CAGGGGCATT AAGAAAACAA ATTGAATCAG GCGCACCTGT TGACGTATTT    4800

ATGTCTGCAA ATACTAAAGA TGTAGATGCA TTAAAAGACA AGAATAAAGC GCATGATACA    4860

TATAAATATG CGAAAAATAG TCTAGTATTA ATTGGTGATA AGATTCAAA TTACACTTCA    4920

GTAAAAGACT TAAAAGACAA TGATAAATTA GCATTAGGTG AAGTGAAAAC TGTACCAGCA    4980

GGAAAATATG CGAAACAGTA TTTAGATAAC AATAACTTAT TTAAAGAAGT CGAAAGTAAA    5040
```

```
CAAGGTTTTG TGTATAAAAC TGACTTATAT AAACAAAATA AAAAAATTGA TACTGTAAAA        5160

GTAATTAAAG AAGTAGAACT TAAGAAGCCA ATCACATACG AAGCTGGTGC TACATCAGAT        5220

AGTAAATTAG CAAAAGAGTG GATGGAATTC TTAAAATCAG ATAAAGCTAA AGAAATACTA        5280

AAAGAATACC ACTTTGCAGC ATAAGGAGTT GTAATCCATG CCTGACTTAA CACCTTTTTG        5340

GATATCAATA CGAGTTGCTG TAATCAGTAC GATTATTGTA ACGGTTTTAG GTATTTTTAT        5400

ATCTAAATGG TTGTATCGTC GTAAGGGTTC GTGGGTTAAA GTATTGGAAA GTTTATTGAT        5460

ATTACCTATT GTTTTGCCGC CAACGGTATT AGGTTTTATT CTATTAATCA TCTTCTCGCC        5520

AAGAGGACCA ATCGGTCAAT TCTTTGCGAA TGTACTACAT TTACCTGTAG TGTTCACTTT        5580

GACAGGTGCT GTGATAGCAT CTGTCATTGT TAGTTTTCCA CTAATGTATC AACATACTGT        5640

GCAAGGCTTC AGAGGTATAG ACACGAAAAT GATTAATACA GCTAGAACGA TGGGAGCAAG        5700

TGAAACGAAA ATTTTCCTCA AATTAATTTT ACCATTAGCT AAACGCTCTA TTTTAGCAGG        5760

TATAATGATG AGTTTTGCTC GTGCATTAGG TGAGTTTGGT GCTACATTAA TGGTTGCAGG        5820

ATATATTCCA AATAAAACGA ATACACTACC TTTAGAAATA TACTTCTTAG TGGAACAAGG        5880

TAGAGAAAAT GAAGCGTGGT TATGGGTATT AGTGCTAGTC GCATTCTCTA TTGTGGTTAT        5940

ATCTACAATT AATTTATTGA ATAAAGATAA ATATAAGGAG GTCGACTAGA TGCTTAAAAT        6000

CAATGTGAAA TATCAATTAA AGAACACTTT AATTCGCATC AATATAGATG ATACTGAACC        6060

AAAAATTTAT GCAGTTCGTG GTCCATCTGG CATTGGTAAA ACTACTGTTT TAAATATGAT        6120

TGCCGGATTA CGTAAAGCAG ATGAAGCTAT TATCGAAGTG AATGGGCAAT TACTTACTGA        6180

TACGGCAAAA AACGTGAATG TTAAAATTCA ACAACGACGT ATTGGATATC TGTTTCAAGA        6240

CTACCAATTG TTTCCTAATA TGACGGTCTA TAAAAATATT ACTTTTATGG CTGAACCATC        6300

TGAACACATC GATCAATTAA TTCAAACTTT AAACATTGAT CATTTGATGA AACAATATCC        6360

TATGACATTG TCAGGTGGAG AGGCACAACG TGTAGCACTT GCACGTGCAC TTAGCACrAA        6420

ACCAGATTTA ATTTTATTAG ATGAACCTTT TTCTAGTTTG GATGATACTA CAAAAGATGA        6480

GAGTATTACA TTAGTTAAAC GTATTTTCAA CGAATGGCAA ATACCAATCA TATTTGTGAC        6540

ACATTCAAAC TATGAAGCAG AACAAATGGC TCATGAAATT ATTACAATTG GGTAATCATT        6600

TATTTGCCAT TAAAGAGTTT AGAACGTATT TAAAATTGTA GAAGTGAATG CTTCTATCAG        6660

CATTTTAATG ATGTTTTAAA CTCTTTTTTA GGGGCAGTTT TTTTGAGAGA CATTGACGCG        6720

CGTCATATAA TGAAAGTAAT GATAAAAAGA AAGGATAACT TAATGTGAGT CAAGAACGTT        6780

ATTCAAGGCA AATTTTATTT AAACAAATAG GTGAAATAGG TCAAAGCAAA ATAAATCAAA        6840
```

```
GAGCAGGCAT TGCCAAACTA ATCATTGTTG ATAGAGATTA TATTGAATTT AGTAATTTAC    6960

AAAGACAAAC ATTGTTTACT GAAGAAGATG CTTTGAAAAT GATGCCTAAG GTGGTTGCAG    7020

CTAAAAAGCA TTTGCTAGCG TTACGTAGTG ATGTTGATAT TGATGATTAT ATTGCCCATG    7080

TGGATTATTA TTTTTTGGAA ACACATGGAC AGGACGTTGA CGTTATTATT GATGCAACCG    7140

ATAACTTTGA AACACGACAA CTGATTAATG ATTTTGCATA TAAATATCGT ATACCTTGGA    7200

TTTATGGTGG TGTTGTACAG AGTACATATA CAGAAGCTGC ATTTATACCT GGTAAAACAC    7260

CTTGCTTTAA CTGTTTGGTA CCACAATTGC CAGCATTAAA TTTAACATGT GATACAGTAG    7320

GGGTCATTCA ACCTGCCGTG ACGATGGCAA CAAGTTTACA ATTAAGAGAT GCGATGAAAG    7380

TATTAACGGA ACAACCAATT GACACAAAAA TAACTTATGG CGATATTTGG GAAGGTAGTC    7440

ATTATTCATT TGGTTTCAGT AAAAATGCAAC GTTCAGACTG TACAACTTGT GGAGATGTAC    7500

CAAGTTATCC GTATTTAAAC AAGAATGAAC AACGTTATGC AACATTGTGT GGTAGAGACA    7560

CTGTACAGTA TGAAAATGCA TCAATTACAC ACGACATTCT TGTTCAATTT TTAAAACAAC    7620

ATCAGTTAAA TTATCGCAGT AATTCGTATA TGGTTATGTT TGAATTTAAA GGACACCGCA    7680

TTGTTGCTTT TAAAGGTGGA AGGTTTTTAA TACATGGCAT GACACGCACA TCAGATGCCA    7740

CACATCTAAT GAATTTATTG TTTGGATAAA AAAAGATAAG ACAAAAGGAG TGTAATATTA    7800

TGGGCGAACA TCAAAACGTT AAAATTGAATC GTACAGTTAA AGCAGCCGTA CTAACGGTAT    7860

CAGATACTAG AGACTTTGAT ACAGATAAAG GTGGTCAATG CGTGCGCCAA CTATTACAAG    7920

CAGATGACGT TGAAGTGAGT GACGCACATT ATACAATTGT GAAAGATGAA AAAGTAGCCA    7980

TCACGACGCA GGTGAAGAAG TGGTTAGAAG AAGATATTGA TGTCATCATT ACGACTGGTG    8040

GAACAGGTAT TGCACAACGT GATGTGACGA TTGAAGCAGT AAAACCACTT TTAACTAAAG    8100

AGATAGAAGG CTTTGGGGAA TTGTTTAGAT ATTTGAGTTA TGTTGAAGAT GTTGGCACGC    8160

GTGCATTATT GTCTCGTGCT GTAGCAGGTA CAGTTAATAA TAAATTGATA TTTTCGATTC    8220

CAGGATCAAC AGGCGCAGTT AAAATTAGCAT TAGAAAAGCT CATTAAACCA GAATTAAATC    8280

ATCTGATTCA TGAGCTTACA AAATAATTTA TTGATTTGAT TGGCGTTGAA AATCTCCAGA    8340

TTTACCGCCA GACTTGCTTT CAAGGTAGGT TTCGCCAATA ATCATACCTT TATCAACTGC    8400

TTTCGTCATG TCGTAAATGG TTAAAGCCGT TGCTGATGCA GCGGTTAAAG CTTCCATTTC    8460

AACACCGGTT TTGCCAGTTG TAGAGACAGT TGTTTGAATG TTTAAAGTAT AAAGGGGTGC    8520

ATTTGTTTCA TCCCAGCTGA AGTGAACATC TATGCCAGTC AATGGTAATG GATGGCACAT    8580

CGGAATAAGT GTTGATGTAT TTTTGGCAGC CATAATACCA GCGATTTGAG CAGTGTTCAA    8640
```

365

```
AATGCTTGAA TGAGCGACAG CAGTTCTTTT TGTAATTTGT TTGTCTGATA CATCGACCAT      8760

TTTGGCGTGG CCTTGTTGAT TAATATGAGT AAACTCAGTC ATTTTACCCC TCCTAGTGCA      8820

TCTAGTATAT CATGAAAAAA TAAAAGTTTT GGAGATGATT TTTAATGGTA GTAGAAAAAA      8880

GAAACCCAAT CCCAGTTAAA GAAGCAATTC AACGTATCGT TAATCAGCAG AGTTCAATGC      8940

CGGCAATTAC GGTAGCACTT GAAAAAAGTC TAAATCATAT CTTAGCAGAA GATATTGTAG      9000

CTACTTATGA TATACCAAGG TTTGATAAAT CACCTTATGA TGGTTTTGCA ATTCGCAGTG      9060

TTGATTCACA AGGGGCAAGT GGTCAGAATC GCATTGAGTT TAAAGTGATT GATCATATTG      9120

GTGCAGGTTC AGTTTCTGAT AAATTAGTTG GGGATCACGA AGCGGTGCGT ATTATGACTG      9180

GAGCACAAAT ACCTAATGGC GCAGATGCTG TTGTTATGTT TGAACAAACG ATTGAACTAG      9240

AAGATACATT TACAATTCGT AAACCATTTT CAAAAAATGA AAATATATCT TTAAAAGGTG      9300

AAGAAACAAA GACAGGCGAT GTTGTTCTAA AAAAAGGACA AGTAATTAAT CCAGGGGCTA      9360

TCGCGGTCCT TGCAACATAT GGCTATGCAG AGGTTAAAGT TATTAAGCAA CCGAGTGTCG      9420

CTGTTATTGC AACAGGAAGC GAATTATTAG ATGTTAATGA TGTATTAGAA GATGGGAAAA      9480

TTCGTAACTC TAATGGCCCA ATGATTCGTG CCTTAGCAGA AAAATTAGGT CTTGAAGTTG      9540

GTATTTACAA AACACAAAAA GATGATTTAG ATAGTGGCAT CCAAGTCGTT AAAGAAGCTA      9600

TGGAAAAACA TGATATCGTT ATTACAACGG GCGGAGTTTC TGTTGGAGAT TTTGACTATT      9660

TACCTGAGAT TTATAAGGCT GTAAAGGCGG AAGTGTTATT TAATAAAGTA GCAATGCGTC      9720

CTGGTAGCGT AACAACGGTT GCATTTGTAG ATGGaAAGTA TTTGTTTGGa TTATCTGGAA      9780

ATCCATCAGC TTGTTTTACA GGATTTGAAC TATTTGTGAA nCCAGCTGTT AAACATATGT      9840

GTGGCGCACT AGAAGTCTTC CCGCAAATAA TTAAAGCAAC ATTAATGGAA GATTTTACCA      9900

AGGCAAACCC ATTCACACGA TTTATACGTG CTAAAGCAAC GTTAACAAGT GCTGGAGCTA      9960

CTGTAGTACC TTCAGGATTC AATAAATCAG GTGCGGTTGT AGCGATTGCA CATGCTAACT     10020

GTATGGTCAT GTTACCAGGA GGGTCACGTG GTTTTAAAGC GGGGCATACA GTAGATATTA     10080

TATTGACTGA ATCTGACGCT GCTGAAGAGG AACTTCTTTT ATGATTTTAC AAATTGTAGG     10140

TTACAAAAAG TCTGGTAAGA CAACATTGAT GAGGCATATT GTCTCTTTCT TAAAGTCACA     10200

TGGTTATACA GTTGCTACTA TTAAACATCA TGGGCATGGT AAGGAAGATA TTCAATTACA     10260

GGATTCAGAC GTCGATCACA TGAAGCATTT TGAAGCGGGG GCAGATCAAA GTATTGTACA     10320

AGGTTTTCAA TATCAGCAAA CTGTAACACG TGTAGATAAT CAAAATCTTA CTCAAATTAT     10380

TGAAAAATCT GTTACAATTG ACACCAATAT CGTATTAGTT GAAGGCTTTA AAAATGCTGA     10440
```

```
GAATGTTTGT TATAGCATTA ATGTAAGGGA GCATGAAGAT TTTACAGCAT TTGAGCAATG    10560

GTTATTAAAT AAAATTAAAA ATGATTGTGA TACACAATTA ACATAGAGGA TTGAAATGAA    10620

TGAAACAATT TGAAATCGTG ACAGAACCGA TACAAACAGA ACAATATCGT GAATTCACTA    10680

TAAATGAATA TCAAGGTGCA GTAGTTGTTT TTACCGGTCA TGTTCGCGAA TGGACTAAAG    10740

GCGTCAAAAC GGAATATTTA GAATATGAAG CGTATATTCC AATGGCTGAA AAGAAATTGG    10800

CACAAATTGG AGATGAAATA AATGAAAAAT GGCCTGGAAC GATAACGAGT ATTGTTCATA    10860

GAATAGGGCC ATTACAAATT TCAGATATCG CTGTATTAAT TGCGGTTTCT TCACCGCATC    10920

GTAAAGATGC CTATCGAGCA AATGAATATG CAATTGAGCG TATAAAAGAA ATTGTTCCGA    10980

TTTGGAAAAA AGAAATTTGG GAAGATGGTT CAAAATGGCA AGGGCATCAA AAAGGGAATT    11040

ATGAAGAAGC AAAGAGGGAG GAATAAGAGA GATGAAGGTA CTTTACTTCG CAGAAATTAA    11100

AGATATATTA CAAAAAGCAC AGGAAGATAT TGTGCTTGAA CAAGCATTGA CTGTACAACA    11160

ATTTGAAGAT TTATTGTTTG AACGTTATCC GCAAATCAAT AATAAAAAGT TTCAAGTTGC    11220

TGTAAATGAG GAATTTGTAC AAAAATCGGA TTTCATTCAA CCTAATGATA CTGTTGCATT    11280

AATTCCACCG GTTAGTGGAG GTTAAGGGAG CATGAAAGCA ATAATTCTTG CAGGTGGTCA    11340

TTCAGTGCGA TTTGGTAAGC CCAAAGCTTT TGCGGAAGTG AACGGTGAGA CCTTTTATAG    11400

TAGAGTAATT AAGACATTAG AATCAACAAA TATGTTCAAT GAAATTATTA TTAGTACAAA    11460

TGCGCAATTG GCAACGCAAT TTAAATATCC AAATGTTGTT ATAGATGATG AGAATCATAA    11520

TGATAAAGGT CCATTAGCAG GAATTTATAC AATCATGAAG CAACATCCTG AAGAAGAATT    11580

GTTTTTTGTC GTTTCTGTTG ATACACCAAT GATTACTGGT AAAGCTGTAA GCACGTTGTA    11640

TCAGTTTTTA GTTTCTCATC TTATTGAAAA TCATTAGAT GTCGCAGCTT TTAAAGAAGA     11700

TGGACGTTTT ATTCCAACAA TTGCATTTTA TAGTCCGAAT GCATTAGGCG CTATAACTAA    11760

AGCACTACAT TCTGATAATT ACAGTTTTAA AAATGTATAT CATGAATTAT CAACGGATTA    11820

TTTGGATGTA AGGGATGTAG ATGCGCCCTC ATATTGGTAC AAAAATATAA ATTATCAGCA    11880

TGATTTGGAC GCTTTAATTC AAAAATTGTA AGCTGTTAGG AGGTCCACAA ATGGTAGAAC    11940

AAATAAAAGA TAAACTAGGA CGTCCCATCC GTGACTTACG GTTATCTGTG ACAGATCGGT    12000

GTAACTTTAG GTGTGATTAT TGCATGCCTA AAGAGGTATT TGGAGATGAT TTCGTATTTT    12060

TACCTAAAAA TGAACTTTTA ACGTTTGATG AAATGGCTAG AATCGCTAAG GTATATGCAG    12120

AATTAGGTGT AAAAAAAATA CGCATTACAG GTGGAGAACC ATTGATGCGA CGGGATTTAG    12180

ATGTACTTAT AGCTAAATTA AATCAAATCG ATGGTATTGA AGATATTGGT TTGACTACAA    12240
```

```
ATGTCAGTTT GGATGCTATT GATGATACGC TATTTCAATC AATCAATAAT CGTAATATTA        12360

AAGCGACTAC GATTTTAGAA CAAATTGATT ACGCGACGTC TATTGGTTTG AATGTAAAAG        12420

TAAATGTTGT TATACAAAAA GGTATTAACG ATGATCAAAT CATACCAATG CTTGAATATT        12480

TTAAAGATAA ACATATAGAG ATTCGATTTA TAGAATTTAT GGATGTTGGT AATGATAATG        12540

GATGGGATTT CAGTAAAGTT GTAACTAAAG ATGAAATGCT TACAATGATA GAGCAGCACT        12600

TTGAAATCGA TCCTGTAGAA CCAAAATATT TTGGGGAAGT AGCAAAATAT TATCGCCATA        12660

AGGATAATGG TGTTCAATTT GGTTTGATTA CAAGTGTTTC ACAATCATTT TGTTCTACAT        12720

GTACACGCGC AAGGCTGTCA TCAGATGGGA AGTTTTACGG ATGTTTATTT GCAACTGTCG        12780

ATGGATTTAA CGTTAAAGCG TTTATTCGTT CTGGCGTGAC CGACGAAGAA TTAAAAGAAC        12840

AATTTAAAGC TTTATGGCAA ATAAGAGATG ATCGATATTC AGATGAGAGA ACTGCTCAAA        12900

CAGTTGCCAA TCGTCAACGT AAAAAGATAA ACATGAATTA TATTGGTGGT TAATGTGTAG        12960

GGACCACTAC ATATTAAATC ATTAGAGATG TTTTAATATT TCTGTCTTAC TCCCTAAAAT        13020

ACAATATTAT TTATTAAAGT AAAAACGGTC ATATCTATGC CAGATTTAAT AGAAATGATC        13080

GTTTTTAAAG TTTTTACAAG TTGGCGGGGC CCCAACACAG AAGCTGACAG AAAGTCAGCT        13140

TACAATAATG TGCAAGTTGG CGGGGCCCCA ACATAGAGAA TTTCAAAAAG AAATTCTACA        13200

GACAATGCAA GTTGGGGAAC GGGGCCCCAA CACAGAAGGT GACGAAAAGT CAGCATACAA        13260

TAATGTGCAA GTTGGCGGGG CCCCAACATA GAGAATTTCA AAAGAAATTC TACAGACAAT        13320

GCAAGTTGGG GATCAACGAA ATAAATTTTA TGAGAATATC ATTTCTATCC CACTCTTAAG        13380

AATCACTACA TAATAAATCT TTAGTGGTTC TTTAACATTG ATGTCACACT CCATGCCATT        13440

GAGTTGTAAT ATATCTTTTT TAGGTATAAA TGTTGTCGAA TAAACAACAA GTTGTCCAAA        13500

AGATATAAAT CTAAACAAGA TATAGCCAGC AATTTAATAT TTGTAATAGA TAAAATGCTA        13560

AGTTTGATAT ATAATAAATT TAAGTAATTG TATAATAATA TGAATTACAA ACATCTAAGA        13620

AGAAACATAG GAGGCATCAT ATTATGAGTA ATAAAGTTCA ACGTTTTATA GAAGCAGAAA        13680

GGGAGTTAAG TCAGTTAAAG CACTGGTTAA AAACAACACA TAAGATTTCA ATTGAAGAAT        13740

TTGTAGTCCT TTTTAAAGTG TATGAAGCTG AAAAGATTAG CGGTAAAGAA TTGAGGGATm        13800

CATTACATTT TGAAATGCTA TGGGATACAA GTAAAATCGA TGTGATTATC CGTAAAaTCT        13860

ATAAAAAAGA GCTTATTTCT AAATTGCGTT CTGAAACGGA TGAAAGACAA GTATTCTATT        13920

TCTATAGTAC TTCTCAAAAG AAATTGTTAG ATAAAATTAC TAAAGAAATA GAAGTGTTAA        13980

GCGTTACAAA CTAAAAACTT aAAAAgcaTG CCAATCTCTA TTCATCATAA TTGCGTCTTG        14040
```

```
GTTCATGGCA TTTCTAGTTA CATGACGTCC ATGAATTAAG AAGTAAACAA GCATAGTAAT      14160

GATTGCTAAA GCGGCCATAA AGCCGAAGAT TTCACTATAT GAAAACATAT GAGTAAATAA      14220

CCCAAGGAAT GATGGACCGA AGCCGACACC TGCATCTAGA CCAACGTAAA AAGTAGATGT      14280

CGCGATACCA TATTTAATCG GGGGTGAGAC TTTTATCGCA ATAGATTGCA TTGCAGATGA      14340

TAAATTTCCA TACCCTAAAC CTAGGCAAGC ACCAGCAAGT AATATTAACC AGCTTTGATA      14400

GCTTGAAATT AAGCATACAA ATGAAAGGAA AAGCATGATA AATGCTGGGT AGACAATAAT      14460

ATTTTCATTT TTATCATCCA TCAATCTACC AGCAATAGGT CTAGTAATTA ACGATGCTAT      14520

AGCATAGCAA ATAAAGAAAT AGCTTGCTGC AGTGACTAGG TGTCGCTCTA AAGCAAATGC      14580

TTGTAAATAA GTTAGGATGG ACGCATAGGT AACGCCAATT AAAAGCATAA TTACAGCAAC      14640

AGGAATGGCC TCTTTTGCAA TAAATTGATG AATACTAAAT CTTGGTTTAT CAATGACATT      14700

AGTTTCAGTT TTGTTATTTG TTACTTCGAA ATCAACTTTT ATAAATAATG AGATAATGAG      14760

TCCGAGTATG CCTAATATGA CACAAATAAT AAACAGTAAG TCAATTGCGT ATTTTGTAAT      14820

AAGTAACATG CCTAGAAATG GGCCAATCGC TGTACCTAAT ACTAAACTTA AGGAAAATAA      14880

ACTGATGCCT TCACTTTTTC TATTAACAGG GGTAACGTAT GCCGCAATAG TACCTGTTGC      14940

AGTTGTCACA ACTGCAGTTG CGATACCGTT TATGAGACGT ACAAAGATTA AAAAAGCTAA      15000

AGATCCATCA ATAAAATAAA GTAATTGCGT GATAATTAAA GCAATTAAAC CAATAAATAA      15060

TAATCGTTTA GGTCCrATTT sATTTACAAA TTTACCTGTA GCAAATCGA                  15109
```

(2) INFORMATION FOR SEQ ID NO: 45:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 9072 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```
GAGAGTCAAT GGCAAGAAGA ATATAAATAT TTGAGAGCGT TAATCTTTAA TGAAACAGAA       60

TTAGAGGAAG CGTATAAATG GATGCATCCT TGTTACACGT TGAATAATAA AAATGTAGTA      120

CTTATCCATG GCTTCAAAAA TTATGTTGCA CTATTATTTC ATAAAGGTGC CATTTTGGAG      180

GATAAATATC ATACACTCAT TCAACAGACT GAAAAGGTGC AAGCAGCTCG TCAGTTACGA      240

TTTGAAAATT TAACAGAGAT TCAAGCACGT ACCGAAGAAA TTAAATATTA TCTAGCCGAA      300

GCAATTAAAG CTGAAAAAGC TGGTAAAAAA GTTGAAATGA AGAAAACAGA GGAATATGTT      360
```

AAATTAACGC CAGGCAGACA ACATCAATAT ATATATCATA TTGGACAAGC TAAACGCAgT   480

GgAACAAGAC AAAAGCGTGT TGAAAAGTAT ATTAACCAAA TACTAGAAGG TAAAGGGATG   540

CATGATAAGT AATTAATGAG TAAAGCATAC CGGTTATACA ACAACATACA AGATGACACG   600

AAACAACCAA TGGCTCATGC TGTTGGTTGT TTTTTTAGGT GTGTCTGTCA TGGGCAACAC   660

TTTGACGTTG GAATTCCGTT ACAGGCTTGG GAGTAGAAAA TGTTAGCAAA AGGCAAGGGT   720

GTCTACAATG AATGATGAAG ATATTAAAAT ATAAGGATGA CTTTGTGAGT GGCGGATGGG   780

CGGTTGTCCG TCTGTAACAA TGGATGCGTG TGCATTATTA CAAAAATTCG ACTTTTGTAA   840

TAATATTTCA CATTTTCGAC ACTTTTTTGC TATAAAACAA CCAATTGAGC GATAATAAAT   900

TCGCTTTTAA AAAATATGAG TTATCTATTT AGTTGCCAAA GATAAAATAA TAATGTTTAA   960

TAACATCATA TAGAGTATGT TAGTTTTAAA TGTCGAATAT ACGAATGTGc AAACAAAGTA   1020

ATCGGTAGAA ATTCAACATA CATAGCGCCG TTTACTGTTA AGTATTCACA TTACAGATGA   1080

AAAATATAAA ATTCTACATA ATCAAGACCA TGATGTGTAC TTGTTTAACT TATGACTCTA   1140

TTTGTTTAAC AATTGCGATA ATGGTCTTTT TATTTTATGC GTATCATTCG TCATATTTTT   1200

TATGAGGAAG GAGAAATGAT TATGTTAAGT ATTAAGCATT TAACGAAAAT TTATTCTGGT   1260

AATAAAAAGG CAGTAGATGA CATCTCTTTA GATATTCAAT CTGGGGAATT TATCGCATTT   1320

·ATTGGAACCA GTGGAAGTGG CAAAACGACT GCTTTAAGAA TGATAAACCG TATGATTGAA   1380

GCGACAGAAG GACAAATTGA AATTGATGGT AAAGATGTTC GGAGTATGAA TCCTGTCGAA   1440

TTGCGTAGAA ATATTGGCTA TGTTATTCAA CAAATTGGCT TAATGCCTCA TATGACGATT   1500

AAAGAGAATA TTGTGTTGGT ACCCAAATTG TTGAAATGGA CTAAAGAGGA AAAGGATAAA   1560

CGTGCAAAGG AATTAATTAA ACTTGTGGAT TTACCGGAGT CATTTTTAGA GCGTTATCCA   1620

GCAGAACTAT CAGGTGGGCA ACAACAACGT ATCGGTGTTG TAAGAGCACT TGCGGCCGAA   1680

CAAGATATTA TTTTAATGGA TGAACCTTTT GGTGCATTGG ATCCTATTAC GAGAGATACG   1740

TTACAAGATT TAGTTAAAAC GTTACAACGA AAATTAGGCA AGACGTTTAT CTTTGTAACA   1800

CATGATATGG ATGAAGCGAT TAAATTAGCA GACAAAATTT GTATTATGTC AGAAGGTAAG   1860

GTGGTGCAAT TTGATACGCC AGACAATATT TTAAGACATC CCGCAAATGA TTTTGTACGT   1920

GATTTTATAG GACAAAATAG ACTGATTCAA GACCGTCCCA ATGACAAGAC TGTAGAAGGT   1980

GTAATGATTA AACCAATCAC GATACAAGCA GAAGCAACAC TGAATGACGC CGTTCATATT   2040

ATGAGACAAA AACGTGTTGA TACTATTTTT GTAGTAGATA GTAATAACCA TTTACTAGGT   2100

TTCTTAGACA TTGAAGATAT AAATCAGGGT ATACGTGGAC ACAAAAGTTT ACGAGACACC   2160

```
ATTTTAAAAA GAAACGTTAG GAATGTACCT GTCGTAGATG ATCAACAGCG TTTAGTAGGA    2280

CTGATTACGC GTGCCAATGT TGTTGATATT GTATATGACA CGATTTGGGG CGATAGTGAG    2340

GATACAGTGC AAACAGAACA TGTGGGGGAA GACAcTGCGT CCTCAAAAGT GCATGAGCAA    2400

CACACTACTA ATGTCAAAGT ACGTGACATA GGAGATGATA AATCATGATT GAGTTCCTAC    2460

ATGAACATGG TGGACAGTTG ATGTCGAAAA CACTGGAACA TTTCTATATT TCTATAGTGG    2520

CATTATTACT TGCCATCATT GTTGCAGTAC CTATAGGCAT TTTATTATCA AAAACAAAGC    2580

GAACTGCCAA TATTGTATTA ACTGTGGCAG GTGTCTTACA AACTATTCCA ACACTAGCTG    2640

TACTTGCTAT TATGATACCG ATTTTTGGTG TTGGTAAAAC GCCTGCAATT GTAGCGCTAT    2700

TTATTTATGT ATTATTACCT ATTTTAAATA ACACGGTACT CGGTGTTCAA AATATTGATA    2760

GCAACATTAA AGAAGCTGGA AAAAGTATGG GAATGACACA ATTTCAATTG ATGAAGGATG    2820

TTGAATTGCC GTTAGCATTG CCGCTTATCA TTGGTGGCAT TCGTTTGTCA TCTGTGTATG    2880

TAATTAGTTG GGCTACACTT GCAAGTTATG TAGGTGCGGG TGGATTAGGT GATTTCATTT    2940

TCAATGGTTT AAATTTATAT GATCCACTGA TGATTGTAAC TGCAACGGTA CTCGTTACTG    3000

CACTAGCATT AGGTGTTGAT GCCTTATTAG CTTTAGTTGA AAAATGGGTA GTTCCCAAAG    3060

GCTTAAAAGT ATCTGGATAA TTAGGAGGCT AAGATAATGA AGAAAATTAA ATATATACTT    3120

GTCGTGTTTG TCTTATCGCT TACCGTATTA TCTGGATGTA GTTTGCCCGG ACTAGGTAGT    3180

AAGAGCACGA AAAATGATGT CAAAATTACA GCATTATCAA CAAGCGAATC GCAAATTATT    3240

TCACATATGT TACGGTTGTT AATAGAGCAT GATACACACG GTAAGATAAA GCCAACATTA    3300

GTAAATAATT TAGGGTCAAG TACGATTCAA CATAATGCCT TAATTAATGG GGATGCTAAT    3360

ATATCAGGTG TTAGATATAA TGGCACAGAT TTAACGGGAG CTTTGAAGGA AGCACCAATT    3420

AAAATCCTA AGAAAGCAAT GATAGCAACA CAACAAGGAT TTAAAAAGAA ATTTGATCAA    3480

ACGTTTTTTG ATTCGTATGG TTTTGCGAAT ACGTATGCAT TCATGGTAAC GAAGGAAACC    3540

GCTAAAAAAT ATCATTTAGA GACAGTTTCA GATTTAGCAA AGCATAGTAA AGATTTACGT    3600

TTAGGTATGG ATAGTTCATG GATGAATCGT AAAGGCGATG GCTATGAAGG ATTTAAAAAA    3660

GAGTATGGTT TTGACTTTGG TACAGTGAGA CCAATGCAAA TAGGTCTAGT CTACGACGCA    3720

TTAAACTCAG AGAAGTTAGA CGTTGCATTA GGTTATTCTA CAGATGGTCG AATTGCGGCG    3780

TATGATTTGA AAGTACTTAA AGATGATAAA CAATTTTTCC CACCTTATGC TGCGAGTGCT    3840

GTTGCAACAA ATGAATTATT ACGGCAACAC CCAGAACTTA AAACGACGAT TAATAAGTTG    3900

ACAGGAAAGA TTTCGACTTC AGAGATGCAA CGCTTGAATT ATGAAGCGGA TGGTAAAGGT    3960
```

```
AAAGGTGGTC ATAAGTAATG GAAGGTAATT TATTACAGCA ATTATTCAAT TATTATGTTA      4080

CGAACTTTGG TTATCTATGG GATTTATTTT TCAAACACTT ATTAATGTCT GTCTATGGTG      4140

TGCTGTTTGC AgCTTTAATT GGTATTCCAT TGGGAATCTT GCTTGCaAGA TACACAAAAC      4200

TTTCTGGATT TGTAATTACA ATTGCAAATA TAATTCAAAC AGTTCCAGTC ATTGCAATGT      4260

TAGCTATTTT AATGTTAGTC ATGGGCTTAG GTTCAGAAAC AGTAGTTTTA ACAGTGTTTT      4320

TATATGCGTT ACTTCCAATT ATAAAAAACA CTTATACTGG TATAGCTAGT GTTGATGCGA      4380

ATATTAAGGA TGCTGGCAAA GGTATGGGAA TGACACGCAA TCAAGTGCTA CGAATGATTG      4440

AATTACCGTT ATCTGTTTCG GTTATTATCG GTGGCATTCG TATTGCCTTG GTTGTTGCGA      4500

TAGGTGTTGT TGCCGTTGGA TCATTTATAG GAGCACCTAC GCTTGGTGAC ATTGTGATTC      4560

GTGGTACAAA TGCGACGGAT GGCACAACGT TTATTTTAGC AGGTGCGATT CCGATTGCTA      4620

TCATTGCAAT CGTCATTGAT GTACTATTAA GATTTTTAGA AAAACGATTA GACCCAACAA      4680

CACGACATCG TAAAAATCAA TCTAATCATC GGCCGCAAAG TATTAATATG TAATAGTAGA      4740

AGATGTTTAT AATTTAGCGA TTTCGTTTCA TGATTTATAA AAAATGAGGC TACTCAAGGA      4800

GCTCAAATAA TCTTTGAGTA GCCTTTTTAT AGGTTGTGTT TGTATGCGTT TACACTAAAA      4860

TAGCAATTAT TATCATGAAA GTTTTTGGAT AAAAAGCGTT AATTATTGTA AAAATACTAA      4920

AAAATGAGAT GTTTTATTTA TAATTTTCTG CAAATTTATG ATATTGTTTC TTAATATATC      4980

ATATTAAAAA TTTGTTTTTC TTAAACATAG GAGGCTTATC TAATTCATGG ACACATCAAA      5040

ACAATTTAGA GGTGACAACC GATTGCTTTT GGGTATCGTT TTAGGGGTTA TTACCTTTTG      5100

GCTATTCGCG CAGTCACTTG TTAATCTTGT TGTCCCATTA CAATCAACAT ATAGTAGTGA      5160

CGTTGGAACG ATAAATATCG CTGTTAGCTT ATCTGCCTTA TTTGCTGGTT TGTTTATCGT      5220

AGGTGCTGGT GATGTTGCTG ATAAATTTGG TCGCGTCAAA ATTACTTATG TAGGATTGAT      5280

ATTAAATGTT GTAGGTTCAT TACTCATCAT CATTACACCT TTGCCAGCAT TTTTAATTAT      5340

AGGTAGAATA ATTCAAGGTT TGTCTGCAGC ATGTATTATG CCATCAACAC TTGCTATTAT      5400

TAACGAATAT TATATTGGTA CAAGAAGACA ACGTGCCTTA AGCTATTGGT CTATTGGTTC      5460

TTGGGGTGGT AGTGGTATTT GTACGTTGTT TGGTGGCTTA ATGGCTACAT ATATAGGTTG      5520

GCGTTCAATA TTTGTTGTTT CAATTCTATT AACATTATTA GCAATGTACT TAATCAAACA      5580

TGCACCTGAG ACTAAAGCAG AACCAATCAA AGGTATGAAA GCAGAAGCTA AAAGTTTGA      5640

CGTTATTGGT TTAGTCATTT TAGTAGTGAC GATGTTAAGT TTAAATGTAA TCATCACACA      5700

GACGTCTCAT TTTGGTTTAG TTTCACCGTT AATTCTAGGT TTAATTGTTG TGTTTATCTG      5760
```

AATTTTTAAA AATAGAGGAT ACAGTGGTGC AACTATTTCA AACTTCTTAT TAAATGGTGT   5880

AGCAGGTGGT GCACTTATCG TTATTAACAC GTATTATCAA CAACAATTAG GATTTAATTC   5940

TTCGCAAACG GGTTATATTT CATTAACGTA TTTAATAACA GTGTTGTCAA TGATTCGTGT   6000

AGGTGAAAAG ATTTTATCTC AACATGGTCC GAAGCGCCCA CTATTACTAG GAAGTGGCTT   6060

TACAGTGATT GGGTTAATCT TATTGTCGTT AACATTTTTA CCAGAAGTGT GGTATATCAT   6120

ATCTAGTATA GTTGGATATT TATTGTTTGG TACTGGTTTA GGATTATATG CTACACCATC   6180

AACTGATACA GCAGTTGCTA GTGCGCCAGA TGATAAGTCG GGTGTTGCTT CAGGTGTGTA   6240

TAAAATGGCG TCATCATTAG GAAATGCATT TGGAGTAGCA GTATCTGGTA CGGTTTATAC   6300

TGTGTTAGCA GCTAATTTAA ATTTGAACTT AGGTGGTTTC ACAGGTATGA TGTTTAATGC   6360

CTTGCTAGCA ATTGTTGCAT TTTTAGTCAT TTTACTATTA GTTCCTAAAA ATCAAACGAA   6420

TTTGTAAAAC TGAAATGAAA GCAAGTTATT ATGTAGGGAT TTTAAAGGAA ATTTTGTGAA   6480

AGTAAGTTTA TCATACACAC TTAATGTTGC GTATTGACGT TTAATGTTAG GTGTGTTCTT   6540

TTATAGACGA TAAAAGCTGT GTGCATATTA AGCGAATGAT TTTCAAATTG ACGCTAATAT   6600

GCGAAAGTAG TATTTTTAAA ATGAACAACA ACGATGAAGA GGGGTTTATA GGATGAAAAT   6660

TGCAATTGCT GGATCGGGTG CATTAGGTAG TGGCTTTGGT GCCAAACTAT TTCAAGCAGG   6720

ATATGATGTC ACACTTATTG ACGGATATAC ATCTCATGTT GAAGCGGTTA AGCAACATGG   6780

ATTAAATATA ACGATTAATG GAGAGGCATT CGAGTTAAAC ATTCCGATGT ATCATTTTAA   6840

TGATCAACCG GACGAAAGCA TTTACGATGT TGTCTTTCTA TTTCCAAAGT CTATGCAATT   6900

AAAAGAAGTG ATGGAAGATA TGAAGCCACA TATTGATAAT GAAACGATCG TCGTATGTAC   6960

GATGAATGGT CTGAAGCATG AAGAAGTCAT TGCGCAGTAT GTTGCTCAAT CACAAATTGT   7020

CAGAGGTGTT ACGACTTGGA CGGCAGGTCT TGAAAGCCCT GGACACAGTC ATTTACTTGG   7080

TAGTGGACCA GTTGAAATAG GTGAACTAGT GGATGAAGGT AAAGAAAATG TTATAAAAGT   7140

TGCTGATTTA CTTAACGAAG CGGAATTGAA TGGTGTCATT AGTAAAGATT TATACCAATC   7200

GATTTGGAAA AAGATTTGTG TTAATGGTAC GGCAAATGCA TTAAGCACAG TGTTGGAGTG   7260

TAATATGGCA TCGCTGAATG AAAGTAGTTA TGCGAAGTGT TTGATTTATA AATTAACGCA   7320

AGAAATAGTG CATGTAGCGA CGATTGATAA TGTTCATTTA AATGTTGATG AAGTATTTGA   7380

ATATTTAGTT GATTTAAATG AAaAAGTTGG TGCGCATTAT CCATCCATGT ATCAAGATTT   7440

AATTGTTAAT AATAGAAAAA CTGAAATTGA TTATATTAAT GGCGCAGTTG CAACATTAGG   7500

TAAACAACGT CaTATTGAAG CGCCAGTCAA TCGCTTTATT ACTGATTTAA TTCATACTAA   7560

```
CAATCACGTG ATATTACGGT CATTATTAAG ATTGAAATGT AATAAATAAA GAACAGCAGT      7680

AAGGTACTTT CAAATTGAAA TGATCTTGGT GCTGTTTTTC TTGATTGATC TTCGTCATAA      7740

TTCAGATTTG TCATAGGcTA CGACATACTA TTAGTATTTA CTAGACAGTT TTTACGACGA      7800

CACTTTGAAA AATTTTGAGG CAAATCATTT GGAAGTCTCA CGTGAATTTT·GTAAACTCAT      7860

CAAGCAAGTA ATTATATTAA AAAGACAAAT AGAGAAAAGG TGTTTATAAT GAGTAAAATT      7920

TTTGTAACTG GTGCAACGGG CCTTATTGGC ATTAAATTAG·TTCAAAGACT AAAAGAAGAG      7980

GGGCATGAGG TTGCTGGTTT TACTACATCT GAGAATGGTC AACAAAAGCT AGCTGCTGTT      8040

AATGTAAAAG CATATATTGG TGATATATTA AAAGCTGATA CTATTGATCA AGCGTTAGCA      8100

GATTTTAAAC CAGAAATCAT TATCAATCAA ATTACGGATT TAAAAAATGT TGATATGGCA      8160

GCAAATACGA AAGTACGTAT TGAAGGTTCT AAAAACCTAA TTGATGCGGC GAAAAAGCAT      8220

GACGTTAAGA AAGTAATTGC CCAAAGTATT GCCTTTATGT ATGAACCTGG CGAAGGATTA      8280

GCAAATGAGG AAACTTCACT TGATTTTAAC TCAACTGGCG ATAGAAAAGT AACGGTTGAT      8340

GGTGTGGTTG GTTTAGAAGA AGAAACGGCT CGTATGGATG AATACGTTGT TTTACGTTTT      8400

GGCTGGTTAT ATGGCCCAGG TACTTGGTAC GGAAAAGATG GCATGATTTA TAATCAATTT      8460

ATGGATGGTC AAGTGACACT TTCAGATGGC GTAACATCAT TTGTGCATCT TGATGATGCA      8520

GTTGAAACAT CTATTCAAGC TATTCATTTT GAAAATGGTA TCTATAATGT AGCAGATGAT      8580

GCACCTGTTA AAGGTTCTGA ATTTGCAGAA TGGTATAAAG AACAACTTGG TGTTGAACCA      8640

AATATTGATA TTCAACCTGC GCAACCATTT GAACGTGGCG TAAGCAATGA GAAGTTTAAA      8700

GCGCAAGGTG GTACTCTGAT TTATCAAACT TGGAAAGATG GCATGAATCC AATTAAATAA      8760

TAATTTATCC GTTTAATATA CAAAGAATAA AGACTTGGTC GAATCGTGGA TGATATATTA      8820

TCAAACGCAC GGCTCGAACA AGTCTTTTTT ATTATGTCTT CGTTATCTTT GTATGAAGGA      8880

ATAACAGAAT TACAATTAAT GTACTGAATA ATGCAATTAA TGTTGTGATT AGTGCTAATT      8940

TAATTTCTAT TGGTAGCCAA GTCAGTACAA AAGACCAATT ATTGCTACCG AGAATGAGAT      9000

ATGGTAATGC ATATAATATG AGCGCTAAAG CGATACATAT ACATAATGAT AACCAACTCA      9060

ATACAGCAAT CC                                                          9072
```

(2) INFORMATION FOR SEQ ID NO: 46:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 16826 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

```
GTGGAACAGC TGTAACTATA TCATTTCTTT CAACATTTAT TGGGAAAATG TTAGCTACAT          60

TTCTATATCC GATTAATAAT GTAGTACTTT CATATATnTC TGTAAATGAA AGTGACAATA         120

TAAAGAAGCA ATATTTGaAA ACTAATCTAA TTGCTATAGC TGCCCTATGT TTAGTCATGA         180

TTATATGTTA TCCAATTACA ATAATTATTG TCTCTTTACT GTATAACATT GATTCAAGTT         240

TATATTCGAA GTTTATTATT TTAGGTAATA TAGGTGTTTT ATTCAATGCA GTGAGTATTA         300

TGATCCAAAC TTTAAATACA AAACACGCAT CAATAACATT ACAAGCGAAT TATATGACGC         360

TTCACACGAT TACATTTATA TTCATAACTA TTTTAATGAC AATTGCGTTT GGTCTAAATG         420

GATTCTTTTG GACAACGCTG TTCAGCAACA TTATTAAGTA TGTGATTTTA AATATTATAG         480

GTTTAAAGTC TAAATTCATT AATAAAAAGG ACGTCGATTA GATGAGTGAA AAAAAGATTT         540

TGATTTTATG TCAGTATTTT TATCCGGAAT ATGTATCTTC TGCGACGTTA CCAACTCAAT         600

TGGCGGAAGA TTTAATTGCG AATCACATTA ATGTCGATGT CATGTGTGGA TGGCCATATG         660

AATATAGTAA TCATAAACAG GTTTCTAAAA CCGAGATGCA TCGTGGTATT CGCATTCGAC         720

GTCTCAAGTA TTCGAGGTTT AATAACAAAA GTAAGGTTGG AAGGATCATC AATTTCTTTA         780

GTTTATTTTC AAAATTCGTG ATTAATATAC CTAAAATGTT GAAATATGAT CAGATTCTTG         840

TTTACTCTAA TCCACCAATC TTGCCATTAA TACCAGACGT TTTACACAGA CTGCTTAAGA         900

AAAAATATTC TTTTGTGGTG TATGATATAG CACCTGATAA TGCGATTAAG ACAGGTGCAA         960

CTCGTCCAGG TAGCATGATT GATAAGCTGA TGCGTTACAT TAATAGACAT GTCTACAAGA        1020

ATGCTGAAAA TGTCATTGTC CTTGGTACGG AAATGAAAAA CTACTTACTA AATCATCAAA        1080

TTTCTAAAAA TGCTGACAAT ATCCATGTGA TTCCTAACTG GTATGACATG CGTCAATTAC        1140

AAGACAATCG TATCTATAAT GACACATTTA AAGCTTACCG TGAGCAATAC GACAAAATTT        1200

TATTGTATAG CGGTAATATG GGGCAGTTAC AGGATATGGA GACACTTATC TCATTTTTAA        1260

AATTAAATAA GGATCAGTCT CAAACGTTAA CAATACTTTG TGGTCATGGT AAGAAATTTG        1320

CAGATGTCAA AACGGCAATA GaAGACCATC GTATTGAAAA TGTTAAAATG TTTGAGTTTT        1380

TAACAGGTAC AGACTATGCT GACGTATTAA AAATTGCGGA TGTATGTATT GCATCGCTGA        1440

TTAAAGAAGG CGTCGGTTTA GGCGTGCCGA GCAAGAATTA TGGCTATCTT GCAGCTAAGA        1500

AAGCGTTGGT ACTCATCATG GATAAGCAAT CTGATATCGT TCAACATGTT GAACAATATG        1560

ATGCGGGTAT CCAAATTGAT AATGGCGATG CACATGCCAT TTATAACTTC ATCAACACTC        1620

ACTCGAGTAA GGAATTGCAC GAGATGGGTG AGCGCGCACA TCAACTGTTT AAAGATAAAT        1680
```

```
AAGCGATTAT TCGATGTAGT GAGTTCAATA TATGGTTTAG TAGTTTTAAG TCCGATTCTG      1800

TTAATTACAG CATTACTAAT TAAAATGGAa TCACCTGGAC CAGCCATTTT CAAACAAAAA      1860

AGACCGACGA TTAATAATGA ATTGTTTAAT ATTTATAAGT TTAGATCAAT GAAAATAGAC      1920

ACACCTAATG TTGCAACTGA TTTAATGGAT TCAACATCGT ATATAACAAA GACAGGGAAG      1980

GTCATTCGTA AGACCTCTAT TGATGAATTG CCACAATTAT TGAATGTTTT AAAAGGAGAA      2040

ATGTCAATTG TAGGTCCTAG ACCAGCGCTT TATAATCAAT ACGAATTAAT CGAAAAACGT      2100

ACAAAAGCGA ACGTGCATAC GATTAGACCA GGTGTGACAG GACTAGCTCA AGTGATGGGG      2160

AGAGATGATA TCACTGATGA TCAAAAAGTA GCGTATGATC ATTATTACTT AACACATCAA      2220

TCTATGATGC TTGATATGTA TATCATATAT AAAACAATTA AAAATATCGT TACTTCAGAA      2280

GGTGTGCATC ACTAATGAGA AAAAATATTT TAATTACAGG CGTACATGGA TATATCGGTA      2340

ATGCTTTAAA AGATAAGCTT ATTGAACAAG GACATCAAGT AGATCAAATT AATGTTAGGA      2400

ATCAATTATG GAAGTCGACC TCGTTCAAAG ATTATGATGT TTTAATTCAT ACAGCAGCTT      2460

TGGTTCACAA CAATTCACCT CAAGCAAGGC TATCTGATTA TATGCAAGTG AATATGTTGC      2520

TGACGAAACA ATTGGCACAA AAGGCTAAAG CTGAAGACGT TAAACAATTT ATTTTTATGA      2580

GTACTATGGC AGTTTATGGA AAAGAAGGTC ATGTTGGTAA ATCAGATCAA GTTGATACAC      2640

AAACACCAAT GAACCCTACG ACCAACTATG GTATTTCCAA AAAGTTCGCT GAACAAGCAT      2700

TACAAGAATT GATTAGTGAT TCGTTTAAAG TAGCAATTGT GAGACCACCA ATGATTTATG      2760

GTGCACATTG CCCAGGAAAT TTCCAACGGT TAATGCAATT GTCAAAGCGA TTGCCAATCA      2820

TTCCCAATAT TAACAATCAG CGCAGTGCAT TATATATTAA ACATCTGACA GCATTTATTG      2880

ATCAATTAAT ATCATTAGAA GTGACAGGTG TGTACCATCC TCAAGATAGT TTTTACTTTG      2940

ATACATCGTC AGTAATGTAT GAAATACGTC GCCAATCACA TCGTAAAACG GTATTGATCA      3000

ACATGCCTTC AATGCTAAAT AAGTATTTTA ATAAGTTGTC GGTCTTTAGA AAATTATTCG      3060

GCAATTTAAT ATACAGCAAT ACGTTATATG AAAATAATAA TGCACTTGAA ATTATTCCTG      3120

GAAAAATGTC ACTTGTTATT GCGGACATCA TGGATGAAAC GACAACCAAA GATAAGGCAT      3180

AAGTCATCTA TTAAATAAAA TCAACATACA AATCGTTTTA TTTGGAGGTT ATAGTATGAA      3240

GTTAACAGTA GTTGGCTTAG GTTATATTGG TTTACCAACA TCAATTATGT TTGCAAAACA      3300

TGGcGTCGAT GTGCTTGGTG TTGATATTAA TCAGCAAACG ATTGATAAGT TACAAAGTGG      3360

TCAAATTAGT ATTGAAGAAC CTGGATTACA AGAGGTTTAT GAAGAGGTAC TGTCATCGGG      3420

AAAATTGAAG GTATCTACAA CGCCAGATGC ATCTGATGTT TTTATCATTG CCGTTCCGAC      3480
```

```
TAGTATTTTA TCATTTTTAG AAAAAGGAAA TACCATTATT GTAGAGTCGA CAATTGCGCC      3600

TAAAACGATG GATGATTTTG TAAAACCAGT CATTGAAAAT TTAGGGTTTA CAATAGGTGA      3660

AGATATTTAT TTAGTGCATT GTCCAGAACG TGTACTGCCA GGAAAAATTT TAGAAGAATT      3720

AGTTCATAAC AATCGTATCA TTGGCGGTGT GACTGAAGCT TGTATTGAAG CGGGTAAACG      3780

TGTCTATCGC ACATTCGTTC AGGGAGAAAT GATTGAAACA GATGCACGTA CTGCTGAAAT      3840

GAGTAAGCTA ATGGAAAACA CATATAGAGA CGTGAACATT GCTTTAGCTA ATGAATTAAC      3900

AAAAATTTGC AATAACTTAA ATATTAATGT ATTAGATGTG ATTGAAATGG CAAACAAACA      3960

TCCGCGTGTT AACATCCATC AGCCTGGTCC AGGTGTAGGC GGTCATTGTT TAGCTGTTGA      4020

TCCGTACTTT ATTATTGCTA AAGACCCTGA AAATGCAAAG TTAATTCAAA CTGGACGTGA      4080

AATTAATAAT TCAATGCCGG CCTATGTTGT TGATACAACG AAGCAAATCA TCAAAGTGTT      4140

GAGCGGGAAT AAAGTCACAG TATTTGGTTT AACTTATAAA GGTGATGTTG ATGATATAAG      4200

AGAATCACCA GCATTTGATA TTTATGAGCT ATTAAATCAA GAACCAGACA TAGAAGTATG      4260

TGCTTATGAT CCACATGTTG AATTAGATTT TGTGGAACAT GATATGTCAC ATGCTGTCAA      4320

AGACGCATCG CTAGTATTGA TTTTAAGTGA CCACTCAGAA TTTAAAAATT TATCGGACAG      4380

TCATTTTGAT AAAATGAAGC ATAAAGTGAT TTTTGATACA AAAAATGTTG TGAAATCATC      4440

ATTTGAAGAT GTATCGTATT ATAATTATGG CAATATATTT AATTTTATCG ACAAATAAAA      4500

TGTGTCAAAC TAGGGCATAC ATGATTAAGG AAAGATAAGC TGTCATGTGT TTGAACTTCA      4560

GAGAGGATAA TGTTATGAAA AAAATTATGG TTATTTTCGG TACGAGACCC GAAGCAATAA      4620

AAATGGCACC ATTAGTAAAA GAAATTGATC ATAATGGGAA CTTTGAAGCG AACATTGTGA      4680

TTACAGCACA ACATAGAGAT ATGTTAGATA GTGTGTTAAG TATATTTGAT ATTCAAGCTG      4740

ATCATGATTT AAATATTATG CAAGATCAAC AAACATTAGC AGGCCTTACG GCGAATGCAC      4800

TTGCTAAACT TGATAGCATC ATTAATGAGG AACAACCGGA TATGATTTTA GTACATGGTG      4860

ATACTACAAC GACTTTTGTA GGAAGTTTGG CAGCATTTTA TCATCAAATT CCGGTCGGAC      4920

ATGTAGAAGC TGGACTTCGA ACACATCAGA AATACTCACC ATTTCCTGAA GAGTTAAATC      4980

GAGTCATGGT AAGTAATATT GCTGAATTGA ATTTTGCGCC AACAGTAATT GCAGCTAAAA      5040

ATTTACTTTT TGAAAACAAA GACAAAGAGC GTATCTTTAT TACTGGAAAT ACAGTTATTG      5100

ACGCATTGTC AACAACAGTT CAAAATGATT TTGTTTCAAC GATTATTAAT AAACATAAAG      5160

GCAAGAAAGT TGTTTTACTA ACAGCGCATC GTCGTGAAAA TATTGGGGAA CCGATGCATC      5220

AGATTTTTAA AGCAGTAAGA GATTTGGCAG ATGAATATAA AGATGTTGTC TTCATTTATC      5280
```

```
GGATTGAATT AATTGAGCCA TTAGATGCGA TTGAGTTCCA TAATTTTACA AATCAATCGT   5400

ACCTCGTGCT GACAGATTCT GGTGGTATTC AAGAGGAGGC TCCTACATTT GGAAAACCTG   5460

TGTTGGTATT AAGGAATCAT ACAGAGCGTC CCGAAGGCGT TGAGGCGGGA ACATCGAGAG   5520

TAATTGGCAC AGATTATGAC AATATTGTTC GAAATGTGAA ACAATTGATT GAGGATGATG   5580

AAGCGTATCA ACGTATGAGT CAAGCGAATA ATCCATATGG TGATGGACAA GCATCACGAC   5640

GTATTTGTGA AGCAATAGAA TATTATTTTG GATTGCGCAC AGACAAGCCG GATGAATTCG   5700

TACCTTTACG TCACAAATAA TAAAAAACCC CTAATCATGA AGTTGGTTTA GACAACCAGC   5760

GGTGACTAGG GGTTTTTAAT ATATTTATTT TTGATAGTGG TAGCCAATAT CATATTTGAA   5820

TACTTTATTT GATAATATTG GACTTTGCTG TCCATCGTCA TCACTTTTTA AACGTACATT   5880

TTTATGAGCT TCTTTAAATA CATCGGAATT CAACCAATTA TTAAAGCTAT CTTCAGATTC   5940

CCAAATAGTT AAGATTTTAA CTTCGTCTGT ATCCTCGGTA TTTAATGTTT TAGTGACAAA   6000

CATTTGTTGG AAGCCTTCAA TAGTTTCAAT ACCTTGTCTA TTGTAAAAAC GTTCAATCGT   6060

TTCTTCCGCA CTGCCTTTTT GTAATTGTAA TCTATTTTCT GCCATAAACA TGGGCAATCA   6120

CTCCTCTATT TTATGATTTG ATTTGGGTAA TGTTTTTACA AATGTAAAGA GTACAGCGGT   6180

TTGTATGATA ACCATTATGA TTAATCCTAC ACGGACTGCA AGAACATCCA CCATATAAAT   6240

TGAAAAACCT ATTACAATGT ATAAGCTAAT TAAAATTTTA ATTTTCTGTT GTAGCGTGTA   6300

GCCTCGATGT AAATAAAAGT TTTCTACATA TTCTTTATAA ATTTTTTGAT TAATAAGCCA   6360

ATTGTAAAAG CGATCTGAAC TTCGAGCAAA GCAAAAAACT GCTACGAGTA AAAAAGGGGT   6420

CGTTGGCAGT AAAGGTAATA CGGCACCTGC AATACCAAGC GCTGTAAATA TTAAGCCAAT   6480

GACGATTAAA ATAAGTCGCA TTGAAAAAAC TCCATTCTAG TACTAATGCG CATGTAATAT   6540

TGTTTTAGTA ATATAACTCA TGCTAAATAT AATGTGTATG ATAAGTGCAA TGACTCAGTA   6600

AAATGAAACG ATGTTGAATT ATCCTTGTCA CATTAACGCA TTTTAAGCGC GACTTTCATA   6660

ACAACCAAAC TATTTAATGA GAATTATTCT CAAGTATTAT AGTTATATTA TGTGTTTTAT   6720

TTTTGAAAAG TGCAATATGT TTTCGAAAAT AAGATTATTT TTATGTGCAA AAACGACGCA   6780

AAAGTTTTAA AAATGAGACT TCTGTGAGCT GATTATTTTA TAAAATGTAA ACGCTTACTA   6840

TATAATGTGA ATCATATCGT TTAAAAGCAT TATTAAATAT GATGCTAAGA GATTTATATT   6900

ATAGCCAATA AACAAAGGAG AGATAATATG GCAGTAAACG TTCGAGATTA TATTGCAGAG   6960

AATTATGGTT TATTTATCAA TGGGGAATTT GTTAAAGGTA GCAGTGACGA AACAATCGAA   7020

GTGACTAATC CAGCAACTGG AGAAACACTA TCACATATTA CAAGAGCAAA AGATAAAGAT   7080
```

```
TCAGAACGTG CACAAATGTT GCGTGATATT GGTGATAAAT TAATGGCACA AAAAGATAAA        7200

ATTGCAATGA TTGAAACATT AAATAATGGT AAACCGATTC GTGAGACAAC AGCAATTGAT        7260

ATTCCATTTG CTGCAAGACA TTTCCATTAT TTCGCAAGTG TTATTGAAAC AGAAGAAGGT        7320

ACAGTGAATG ATATCGATAA AGACACAATG AGTATCGTAC GACATGAGCC GATTGGCGTC        7380

GTAGGTGCTG TTGTTGCTTG GAACTTCCCA ATGCTATTAG CTGCATGGAA GATTGCGCCA        7440

gCCATTGCTG CAGGTAATAC AATTGTGATT CAACCTTCGT CTTCAACACC ATTAAGTTTA        7500

TTGGAAGTTG CTAAAATTTT CCAAGAGGTA TTACCTAAAG GTGTTGTCAA TATACTAACG        7560

GGTAAAGGTT CAGAATCAGG TAATGCAATT TTCAATCATG ATGGTGTAGA TAAATTATCA        7620

TTTACGGGCT CAACTGATGT AGGTTATCAA GTTGCCGAAG CTGCAGCAAA ACATCTAGTA        7680

CCCGCTACAT TAGAGCTTGG TGGTAAAAGC GCCAATATCA TATTAGATGA TGCTAATTTA        7740

GACCTTGCAG TTGAAGGTAT TCAGTTAGGT ATTTTATTCA ACCAAGGTGA AGTATGTAGT        7800

GCAGGTTCTC GATTATTAGT TCATGAAAAA ATTTATGATC AATTGGTGCC ACGTTTACAA        7860

GAGGCATTTT CAAATATTAA AGTTGGAAAT CCACAAGATG AAGCTACACA AATGGGTAGT        7920

CAAACTGGTA AGGATCAATT AGATAAAATT CAATCATATA TTGATGCAGC AAAAGAATCA        7980

GATGCACAAA TTTTAGCAGG CGGTCATCGC TTAACTGAAA ATGGATTAGA TAAAGGGTTC        8040

TTCTTTGAGC CGACATTAAT TGctGTGCCA GACAATCATC ACAAATTAGC ACAAGAAGAA        8100

ATATTTGGAC CAGTGTTAAC AGTGATTAAA GTGAAGGACG ATCAAGAAGC AATTGATATA        8160

GCTAATGATT CTGAGTATGG TTTAGCAGGC GGTGTATTTT CTCAAAATAT CACACGTGCA        8220

TTAAATATTG CTAAAGCTGT ACGTACAGGA CGTATTTGGA TTAACACTTA CAACCAAGTA        8280

CCAGAAGGCG CACCATTTGG TGGTTATAAA AAATCAGGTA TCGGTCGAGA AACTTATAAA        8340

GGTGCGTTAA GTAACTATCA ACAAGTTAAA AATATTTATA TTGATACAAG CAATGCTTTA        8400

AAAGGTTTGT ACTAGAATAA ATATCGTTTC TGAAGCGTGT TTGTAGGTCA GTCTAGCGGT        8460

AAGTCTTAAC ATTTAACGGC GTTGTTTAGA TTTTAAGCAA AACAAAATAT ATAGGAACAC        8520

GTATCATGAT ATTAGGATAT AATGACTAAA ATAATAGCAG TAGGATGGTT TTTAATTGCA        8580

AATCATCTTA CTGCTGTTTT TAATTATGCT AATTTGCGAT GCGGCTATTA TAAGGACAGA        8640

GTTGTTTATT AATTATGGTG ATTTAGAAAT ATGAAGTTCA ATATGCAAAG TCATCGTTTG        8700

TTTTAATATG CGGAACAATC ATTAAAGTTA TTGCGATTTT TTGAACTTAA TGAAACTAAA        8760

CAATAAATTT GAGATACTTT TTTGTCATTT TTATGTAACT AACACAATAA TCTCGTACAT        8820

TATTAAAATT TTCTATATGA TAGGAATAAA GCAAAGCGCG AGTGTGCTGT AAAAGTTTTC        8880
```

```
GATGATGTAT AAATCATGGT TAATTACGGA AGCATTAATA TTAACCTGAG AAGCTATAAA    9000

GAATTATTTT TAAAAGCGAC AATATTAAAT ACGACGCATT TATTTAGGAG TGGCAAACGT    9060

ATGAATGGGA AAAAGGCGAA TACGATAAAC AGATACAAAT ATTTTCATCA TGTCAATCAT    9120

CAAAAAATTC AACAAAGTTC TAAAAAGACG CTGTGGGCAT CACTAATCAT CACATTGTTA    9180

TTTACAGTGA TTGAATTTGT CGGAGGTTTA GTATCTAATt CATTGGCATT ACTGTCAGAT    9240

TCATTTCATA TGCTTAGTGA TGTATTAGCA CTTGGTTTAT CTATGTTGGC CATTTATTTT    9300

GCAAGTAAAA AGCCGACTGC ACGATACACA TTTGGATATT TAAGATTTGA GATATTAGCT    9360

GCATTTTTAA ATGGTTTAGC ATTAATTGTA ATTTCAATCT GGATTTTATA TGAAGCTATT    9420

GTACGTATTA TTTATCCGCA ACCAATTGAA AGTGGCATTA TGTTTATGAT TGCTAGTATT    9480

GGTTTACTCG TCAATATTAT TTTGACTGTT ATCCTTGTAA GGTCTTTAAA ACAAGAAGAC    9540

AATATCAATA TTCAAAGTGC ATTATGGCAT TTCATGGGAG ACTTATTGAA CTCTATTGGT    9600

GTCATCGTTG CAGTTGTATT GATTTACTTT ACAGGATGGC GCATCATCGA CCCAATCATT    9660

AGTATTGTAA TTTCACTCAT CATTTTACGT GGTGGTTATA AAATTACGCG TAATGCgTGG    9720

tTAATTTTAA TGGAAAGTGT GCCTCAACAT TTGGATACTG ATCAAATTAT GGCAGATATT    9780

AAAAACATAG ATGGCATATT AGATGTACAT GAATTTCATT TGTGGAGTAT TACAACAGAG    9840

CATTATTCAT TAAGTGCCCA TGTTGTGTTA GATAAAAAAT ATGAGGGTGA TGATTATCAA    9900

GCGATTGATC AAGTATCATC ATTGTTGAAA GAAAAATATG GCATTGCACA TTCAACGTTG    9960

CAAATTGAAA ACTTGCAATT GAATCCATTA GATGAGCCAT ACTTCGACAA ATTAACATAA   10020

ATAAAACATT GTAGCGCCTA AAACATTAAT CTATGTCATA GGCGCACGTT TCGTTTTATA   10080

CTTATGTTGC ATCATTTAAA TGATTTTCGT CAATTTCTTT GATGCTATCT ACATCTAACA   10140

CGACATCTTT AGGTTTCAAA ATATGAATAT GTTTTTCATC ATTTGTATGT AAAATGCGTT   10200

CTATGATGTA CCTTTGACCG GCCATTGTTT CTACAGCAAT CTTTTTGTTT CTAGCTAAAC   10260

TTGCTACGAC AGATTCTTTA TCCATAATGA TAGCCCCCTA TATATATGTT TATTTACTTA   10320

TACCCTAACA TGATTTTTAT ACTCTTTGAA AATATATTTT ACAGAATTTT ATCTAAATAT   10380

TTAAAAAAAT ATCTTAATAT CCTTGTAATC CGATAAGAAT TATAGTAATA TTTTTTCAAC   10440

CATtGTTATA GGAGGTCTTA TTAATGACAT TATTTTTATT AGAAGCTAAC AATCTTGATT   10500

TTGCATCAAC GAAAGAAGAA CTAGAAGCAA AGGCAGCATC ACTATCTACG AAGACAATTC   10560

CAACATTAAT TGAAGTACAA GCTACTGAAA ATTTAACTCA TGGTTATTTT ATTGTGGAAG   10620

CAAATGACGA aGCAGAAGCT AAACAATTTT TAACAGAAGC AGATATTAGT ATTCAATTAG   10680
```

```
     ATATAACAAT TCACGATATA AGGGCTGTGT TTGGCATAGC CCTTTAGATA TACACTTAAT    11280

     TCCTATTAAA ATAGTAGGGA TTAAAAGGGG GCTTGTCATG ATTAAAATTC AACAATTACA    11340
```
```
     ACATCACTTT GGATCACATA AAGTAATTCA TAACTTTAAT TTGGACATTA GCAAGGGAGA    11400

     AATAGTCACT TTCATAGGGA AAAGTGGTTG CGGAAAGTCT ACTTTACTCA ATATTATCGG    11460

     TGGATTTATT CATCCATCGT CTGGTCGTGT CATTATTGAT AACGAAATTA AACAACAGCC    11520
```
```
     ATCTCCAGAT TGTTTAATGC TATTTCAACA TCATAATTTG CTGCCATGGA AAACGATTAA    11580

     TGACAACATT AGGATTGGAT TACAACAGAA AATTAGTGAT GAAGAGATTA ACGCACAGCT    11640

     TAAATTAGTT GATTTAGAAG ACAGGGGAAA GCATTTTCCC GAGCAACTGT CCGGGGGTAT    11700
```
```
     GAAACAACGT GTGGCACTAT GTCGAGCGCA TGTGCATAAG CCTAACGTTA TATTGATGGA    11760

     TGAGCCATTA GGTGCATTAG ATGCATTTAC ACGTTATAAA CTTCAGGATC AACTAGTGCA    11820

     aCTAAAACAT AAAACGCAAT CAACTATTAT TTTAGTGACG CATGACATTG ATGAAGCTAT    11880
```
```
     TTATCTTTCC GACCGCATTG TTCTGTTAGG TGAAGGGTGC AATATTATTT CTCAATATGA    11940

     AATTACAGCA TCACATCCAC GCAGTCGTAA TGATAGCCAC CTACTTAAGA TTCGTAATGA    12000

     AATTATGGAA ACATTTGCAT TGAATCATCA TCAAGTTGAA CCTGAATATT ATTTATAAGG    12060
```
```
     AGTGAGTGAC GATGAAAAGG TTAAGCATAA TCGTCATCAT TGGAATCTTT ATAATTACAG    12120

     GATGTGATTG GCAAAGGACG TCTAAAGAAC GGTCTAAAAA TGCCCAAAAT CAGCAAGTGA    12180

     TTAAAATTGG ATATTTGCCG ATTACACATT CAGCTAATTT GATGATGACT AAAAAATTAT    12240
```
```
     TATCACAATA CAATCATCCG AAATATAAAC TAGAATTAGT TAAATTCAAT AATTGGCCAG    12300

     ATTTAATGGA CGCATTAAAC AGTGGTCGTA TTGATGGTGC ATCAACTTTA ATAGAGCTAG    12360
```
```
     CGATGAAATC AAAACAGAAG GGCTCAAATA TAAAGGCTGT GGCATTGGGC CATCATGAAG    12420

     GCAATGTCAT TATGGGACAA AAAGGTATGC ACTTAAATGA ATTTAATAAT AATGGCGATG    12480
```

381

```
GTAAACAATT AAAGATTAAA CCGGGGCATT TTAGCTATCA TGAAATGTCG CCAGCAGAAA      12600

TGCCAGCCGC ATTGAGTGAA CACAGAATTA CAGGGTATTC TGTAGCCGAA CCATTCGGTG      12660

CACTGGGTGA AAAGTTAGGC AAAGGTAAGA CTTTGAAACA TGGTGATGAC GTTATACCTG      12720

ATGCGTATTG CTGTGTGCTA GTACTGAGAG GGGAATTGCT TGATCAACAC AAGGATGTAG      12780

CGCAAgCATT TGTACAAGAT TATAAAAGT CTGGCTTTAA AATGAATGAT CGCAAGCAAA       12840

GTGTAGACAT TATGACGCAT CATTTTAAAC AAAGTCGTGA CGTTTTAACA CAGTCAGCGG      12900

CATGGACATC CTATGGTGAT TTAACAATTA AGCCATCCGG CTATCAAGAA ATTACGACAT      12960

TGGTAAAACA ACATCATTTG TTTAATCCAC CTGCATATGA TGACTTTGTT GAACCGTCAT      13020

TGTATAAGGA GGCATCGCGT TCATGACACG TCCCACAAAT AACAAATTTA TATTACCTAT      13080

TATCACATTT ATTATTTTCT TAGGCATTTG GGAAATGGTC ATTATTATTG GGCATTACCA      13140

ACCTGTATTG TTACCGGGTC CTGCTCTTGT AGGAAAAAGT ATATGGTCTT TCATTGTTAC      13200

TGGAGAAATT TTCCAACATT TAGCAATTAG TTTATGGAGA TTTGTAGCGG GCTTTGTTGT      13260

CGCATTGTTG GTTGCTATTC CATTGGGCTT CTTGCTTGGA AGGAATCGTT GGCTATACAA      13320

CGCTATCGAA CCGCTATTTC AATTGATTAG GCCGATATCT CCGATAGCAT GGGCACCATT      13380

TGTTGTTCTA TGGTTTGGTA TTGGTAGTTT GCCAGCGATT GCGATTATTT TTATCGCTGC      13440

TTTTTTCCCA ATTGTGTTCA ATACTATTAA AGGCGTTAGA GACATTGAAC CTCAATATTT      13500

AAAAATAGCA GCAAATTTAA ATTTAACTGG GTGGTCATTG TATCGCAATA TATTATTTCC      13560

CGGGGCATTT AAACAAATCA TGGCTGGGAT ACATATGGCG GTAGGAACAA GTTGGATATT      13620

TTTAGTTTCT GGTGAAATGA TTGGTGCACA ATCGGGATTA GGTTTTTTAA TCGTTGATGC      13680

ACGAAATATG TTGAACTTAG AAGATGTTTT AGCAGCAATA TTCTTTATCG..GATTATTTGG      13740

TTTTATTATT GATCGATTCA TTAGTTATAT TGAGCAGTTT ATACTTAGAA GATTTGGTGA      13800

ATAAGGAGAG ATGATGATGA CTTTAGAAAC GCTTATCAAA GAACAATTAG ATCCTCATTT      13860

AGTAGAAGTT GATGAAGGGA CGTATTATCC GAGAACATTT ATTCAGCAAT TATTTGTAGA      13920

TGGTTATTTC GGTGAGGCGG CATTGAGAAA AAATGCTGAA GTAATCGAAG CTGTATCGCA      13980

GTCTTGTTTG ACAACAGGAT TTTGTTTATG GTGCCAATTA GCTTTTTCAA CGTATTTAGA      14040

AAATGCCACG CAGCCACATT TAAATAATGA CTTACAACAG CAATTGTTAT CTGGAGAAAT      14100

ATTAGGTGCT ACCGGATTGT CTAATCCGAT GAAGTCATTT AATGATTTAG AAAAGTTGAA      14160

CCTTGAACAC ACTTATGTTG ATGGACAATT GGTTGTCAGT GGACGTATGC CAGCTGTAAG      14220

TAATATTCAA GAAGACCATT ATTTTGGTGC GATTTCGAAA CATGAATCAT CAGATGAATT      14280
```

382

```
TTTAGGAGTC AACGGGTCAG CAACGTATCA AATCACATTG AATCAAGTCG TAGTGCCACA    14400

ATCACAAATT ATCACGCATG ATGCGAAGCA GTTTGCGGCA ACTATTCGCC CGCAATTTAT    14460

TGCTTACCAA ATTCCAATAG GATTAGGCTC AATTAAAAGT TCTTTAGAGT TAATTGATGC    14520

ATTTTCAAAT GTGCAAAACG GAATAAATCA ATATTTAGAG TATGATGTTG AAGCTTTTAA    14580

AAAACGTTAT CGTCAACTTA GAGAGGAATA TTATGCAATA TTAGATGACG GTAACTTAAC    14640

TTCACATTTA AATGAATTAA TATCATTGAA GAAGGACATC GGCTATTTAT TGTTAGATGT    14700

AAATCAAGCT TCTGTTGTCA ATGGTGGTTC TAGAGCGTAC ACACCATATT CGCCACAAGT    14760

TCGCAAGTTA AAAGAAGGAT TCTTCTTCGC AGCATTGACA CCGACATTAA GACATTTAGG    14820

TAAACTTGAA GCAGAGTTGA AGGGGTAAGT GTGATAAGCT GATTTTTTGT TTAGATGCGT    14880

TTGTTGAAAC ATTTTTTAAA ATAATATAAA TCTTAGTTTA TAAACATTTT CTGTTAATTT    14940

GTTATATCCT TTTAACTAGG AAAATATACA TTTCGTAATA ATAATAATCG TTATCATTGA    15000

AAAAGTGTTA ATAAGGTGTA TAATGAAAAT GTGAACAATT AATGAACTTC TTATTTTAAA    15060

GAAGGTGAAT ACTATAGATA CGCATACTAA AGAACAACAA TTCTCGAATC TAGTAAGATC    15120

TTATCGTAAA GAATACGTGG GTAAAGGACC CAATAGTATT CGAGTGTCGT TTAAAGATAA    15180

TTGGGCGATT GCACATATGA CAGGTGTTTT GAGTAAAGTT GAGAGTTTTT ACCTAAACGA    15240

CAAACGCAAT GAATCGATGC TCCATTATAC ACGCACAGAG AAGATTAAAC AGATGTATAA    15300

AGAAATAGAT GTAAATGAGA TGGAAAGTCT TGTAGGCGCT AAGTTTGTAA AATTATTTAC    15360

AGATATTGAT TTGAATGATG ATGAAGTCAT TTCAATATTT GTTTTCGATA AGTCAATAGA    15420

ATAAGTGTTG CTGGTGTAAG GTACACGGTG CTGTTTGCTA ACTTCGCTTT GAATTTAACA    15480

ATAATTCAAG GGGGTGGTAT GTCAAACGGT GCCGTTTTTT TGTCATATTT TTAAAACAAG    15540

CAACATGCAA CACGTACTTT AAGGAAGTCA AAATTTATCA TTTAGGAGAG ATGGATATGA    15600

AAATCGTAGC ATTATTTCCA GAAGCAGTAG AAGGTCAAGA AAATCAATTA CTTAATACTA    15660

AAAAAGCATT AGGATTAAAA ACATTTTTAG AGGAAAGAGG ACATGAGTTC ATTATATTAG    15720

CAGATAATGG TGAAGACTTA GATAAACATT TACCAGATAT GGATGTGATT ATTAGTGCGC    15780

CATTTTATCC TGCATATATG ACTCGTGAAC GTATTGAAAA AGCACCGAAC TTGAAATTAG    15840

CAATTACAGC AGGTGTAGGA TCTGACCATG TAGATTTAGC GGCAGCAAGT GAACACAATA    15900

TTGGTGTCGT TGAAGTTACA GGAAGTAATA CAGTTAGTGT GGCAGAACAT GCGGTTATGG    15960

ATTTATTAAT ACTTCTTAGA AACTATGAAG AAGGTCATCG TCAATCAGTA GAAGGTGAAT    16020

GGAACTTGTC TCAAGTAGGT AATCATGCGC ATGAATTACA ACACAAAACA ATTGGTATTT    16080
```

TACAACACTA TGATCCAATC AATCAACAAG ACCATAAATT GTCTAAATTT GTAAGCTTTG   16200

ATGAACTTGT TTCAACAAGT GATGCGATTA CAATTCATGC ACCATTAACA CCAGAAACTG   16260

ATAACTTATT TGATAAAGAT GTTTTAAGTC GTATGAAAAA ACACAGTTAT TTAGTGAATA   16320

CTGCACGTGG TAAAATTGTA AATCGCGATG CGTTAGTTGA AGCGTTAgCA TCCGAGCATT   16380

TACAAGGATA TGCTGGTGAT GTTTGGTATC CaCAACCtGC ACCTGCTGAT CATCCATGGA   16440

GAACAATGCC TAGAAATGCT ATGACGGTTC ACTATTCAGG TATGACTTTA GAAGCACAAA   16500

AACGTATTGA AGATGGAGTT AAAGATATTT TAGAGCGTTT CTTCAATCAT GAACCTTTCC   16560

AAGATAAAGA TATTATTGTT GCAAGTGGTC GTATTGCTAG TAAAAGTTAT ACAGCTAAAT   16620

AGAATAAGGA TGCTGGGCTA GCGATTAACG CTTTCAATTT TATATAAATG AATCATATAA   16680

GCACTACTGC TGTTGTAAAG ATGGCAGTAG TTTTTTTATG ATTACATCTA AGTATAGTCA   16740

CGGCTATGTT AGGACAATGA TTTAACATTT ACGCACATAT GTGTTCACTT ACGCAATTAT   16800

TGAnAAATnT CATTCATGTG GnAATC   16826

(2) INFORMATION FOR SEQ ID NO: 47:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 4012 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: double
       (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTCAATGAGA GTAGTGGGCT GATGTTTAGC GATATCGCGT AAGATTAACC ATTGGCCATA   60

ATATATATTG TGTTTTTCTA AAATCGGCTC GGCTAATTTT AAATAGGGGC GATATATTGT   120

TATAAAACTA TTGAAAAATT CTTGTGATAG CATAGTGACA TCTCCTAAGA CAAAATAGTT   180

AGCTTAGCTA mCCTTTTTAC AACAATAGTA ATTATAAAAC GGGAGCAATT AGAAATCAAT   240

ATATAATTAT TAAGAGCAAA AATAATTATA CTTTGTTAAA ATAAGCGTAA TTACATGTAA   300

ATAGGGGGAT ACTAATGATA TTGAAATTTG aTCACATCAT TCATTATATA GATCAGTTAG   360

ATCGGTTTAG TTTTCCAGGA GATGTTATAA AATTACATTC AGGTGGGTAT CATCATAAAT   420

ATGGAACATT CAATAAATTA GGTTATATCA ATGAAAATTA TATTGAGCTA CTAGATGTAG   480

AAAATAATGA AAAGTTGAAA AAGATGGCAA AAACGATAGA mGGCGGAGTC GCTTTTGCTA   540

CTCAAATTGT TCAAGAGAAG TATGAGCAAG GCTTTAAAAA TATTTGTTTG CGTACAAATG   600

ATATAGAGGC AGTTAAAAAT AAACTACAAA GTGAGCAGGT TGAAGTAGTA GGGCCGATTC   660

ATATCAAACA CCTCATTGTT AGATTATTGA CATTATAACA GGGGTAATTG TATATGAACA   1380

TTAATGTGGT TGCTTGAGGA AAAATTTATT CATTGAAGTC AAGTTGGTTC ATTTTAGAAA   1440

TGAATATCGT GTTAGATGAT GAAAGTATAT TGAAGTATAG GTAACTAGTT GAAAAGTATT   1500

AATTGTACGA TAACATTAAA TTTAACACGA AACATAGATA TAAAATGATT CACAATTAAA   1560

ATGGGTAAAT TTGAACTTGC TAAACTATTA ATTGGAGCAT GGACATTTCA AAAATAAGAG   1620

TTCAAATCTT ACACAAGCTC TGAATCGACA CTATAAGATA CAAACTGTAT AATTAAAGGT   1680

ATTGTTAAAT AGAAGGAGAT ATCATAAATC ATGGAAAAGA TGCATATCAC TAATCAGGAA   1740

CATGACGCAT TTGTTAAATC CCACCCAAAT GGAGATTTAT TACAATTAAC GAAATGGGCA   1800

GAAACAAAGA AATTAACTGG ATGGTACGCG CGAAGAATCG CTGTAGGTCG TGACGGTGAA   1860

GTTCAGGGTG TTGCGCAGTT ACTTTTTAAA AAAGTACCTA AATTACCTTA TACGCTATGT   1920

TATATTTCGC GTGGTTTTGT TGTTGATTAT AGTAATAAAG AAGCGTTAAA TGCATTGTTA   1980

GACAGTGCAA AAGAAATTGC TAAAGCTGAG AAAGCGTATG CAATTAAAAT CGATCCTGAT   2040

GTTGAAGTTG ATAAAGGTAC AGATGCTTTG CAAAATTTGA AAGCGCTTGG TTTTAAACAT   2100

AAAGGATTTA AGAAGGTTT ATCAAAAGAC TACATCCAAC CACGTATGAC TATGATTACA   2160

CCAATTGATA AAAATGATGA TGAGTTATTA AATAGTTTTG AACGCCGAAA TCGTTCAAAA   2220

GTGCGCTTGG CTTTAAAGCG AGGTACGACA GTAGAACGAT CTGATAGAGA AGGTTTAAAA   2280

ACATTTGCTG AGTTAATGAA AATCACTGGG GAACGCGATG GCTTCTTAAC GCGTGATATT   2340

AGTTACTTTG AAAATATTTA TGATGCGTTG CATGAAGATG GAGATGCTGA ACTATTTTTA   2400

GTAAAGTTGG ATCCAAAAGA AAATATAGCG AAAGTAAATC AAGAATTGAA TGAACTTCAT   2460

```
CAAAATATGA TTAATGATGC GCAAAATAAA ATTGCTAAAA ATGAAGATTT AAAACGAGAC    2580

CTAGAAGCTT TAGAAAAGGA ACATCCTGAA GGTATTTATC TTTCTGGTGC ACTATTAATG    2640

TTTGCTGGCT CAAAATCATA TTACTTATAT GGTGCGTCTT CTAATGAATT TAGAGATTTT    2700

TTACCAAATC ATCATATGCA GTATACGATG ATGAAGTATG CACGTGAACA TGGTGCAACA    2760

ACTTACGATT TCGGTGGTAC AGATAATGAT CCAGATAAAG ACTCAGAACA TTATGGATTA    2820

TGGGCATTTA AAAAAGTGTG GGGAACATAC TTAAGTGAAA AGATTGGTGA ATTTGATTAT    2880

GTATTGAATC AGCCATTGTA CCAATTAATT GAGCAAGTTA AACCGCGTTT AACAAAAGCT    2940

AAAATTAAAA TATCTCGTAA ATTAAAACGA AAATAGATTA ACGACTGAAA TCTGAACGCT    3000

CATAAGACTG TCATTTGCGT TCAGATTTTT TTACACAATA TAGAATGGTT GAGTAAAATA    3060

TTTTTGAATA TAGTGAAAGA GGGGGAAGTA CTGTGATAAA AAAGCTATTA CAATTTTCTT    3120

TAGGGAATAA GTTTGCTATC TTTTTAATGG TTGTTTTAGT TGTCTTGGGC GGTGTATATG    3180

CGAGTGCTAA ATTGAAATTA GAATTACTAC CAAATGTACA AAATCCAGTT ATTTCAGTTA    3240

CAACAACAAT GCCGGGTGCA ACGCCACAAA GTACCCAAGA TGAAATAAGT AGTAAAATTG    3300

ACAATCAAGT AAGATCATTG GCATATGTGA AAAATGTTAA AACGCAATCC ATACAAAATG    3360

CTTCAATTGT AACAGTTGAA TATGAAAATA ATACAGATAT GGATAAAGCA GAAGAACAGC    3420

TTAAAAAAGA AATCGATAAA ATTAAATTTA AAGATGAAGT TGGTCAACCA GAATTAAGAC    3480

GTAATTCGAT GGATGCTTTT CCGGTTTTAG CATATTCATT TTCAAATAAA GAGAATGACT    3540

TGAAAAAAGT AACGAAAGTA CTGAATGAAC AATTAATACC AAAATTGCAA ACGGTAGATG    3600

GTGTGCAAAA TGCGCAATTA AATGGGCAGA CGAACCGTGA AATCACCCTT AAATTTAAGC    3660

AAAATGAACT TGAAAAATAT GGGTTGACTG CTGATGATGT AGAAACTAT CTAAAAACGG     3720

CAACAAGAAC AACGCCACTT GGATTGTTCC AATTTGGTGA TAAAGATAAT CAATTGTTGT    3780

TGATGGTCAA TATCAATCTG TTGATGCTTT TAAAAACATA AATATTCCAT TAACGTGGCA    3840

GGAGGACCAA GGGCATCTCA TCCCAAAGTG ACCATAAACC AAATTCAGCC ATGTCAGACG    3900

TTATCAGGCA TCACCACAGC AAATTCAAAG CGTCAGCnCC AATATATAGT GGATGCCGCA    3960

nGAACTAGGG GTTTAGCGnT ATCAGTGGTG TGGCGACTCT ATTCTAAACG AT            4012
```

(2) INFORMATION FOR SEQ ID NO: 48:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 7778 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

```
CAATATAGGT CGCCGAGTTT CAACTaCATC AACTGGTTCA GTTACATTAG ATAATGCGCT      60

AGGTGTAGGT GGCTATCCTA AAGGACGAAT TATTGAAATT TATGGTCCTG AAAGTTCTGG     120

TAAGACAACA GTAGCGCTTC ACGCTATTGC TGAAGTACAA AGTAATGGCG GGGTGGCAGC     180

ATTTATCGAT GCTGAACATG CTTTAGATCC AGAATATGCT CAAGCATTAG GCGTAGATAT     240

CGATAATTTA TATTTATCGC AACCGGATCA TGGTGAACAA GGTCTTGAAA TCGCCGAAGC     300

ATTTGTTAGA AGTGGTGCAG TTGATATTGT AGTTGTAGAC TCAGTTGCTG CTTTAACACC     360

TAAAGCTGAA ATTGAAGGAG AAATGGGAGA CACTCACGTT GGTTTACAAG CTCGTTTAAT     420

GTCACAAGCG TTACGTAAAC TTTCAGGTGC TATTTCTAAA TCAAATACAA CTGCTATTTT     480

CATCAACCAA ATTCGTGAAA AAGTTGGTGT TATGTTCGGT AATCCAGAGA CTACACCAGG     540

TGGACGTGCA TTAAAATTCT ATAGTTCAGT AAGACTAGAA GTACGTCGTG CAGAACAGCT     600

TAAACAAGGA CAAGAAATTG TAGGTAATAG AACTAAAATT AAAGTCGTTA AAAATAAAGT     660

GGCACCACCA TTTAGAGTAG CTGAAGTTGA TATTATGTAT GGACAAGGTA TTTCTAAAGA     720

GGGTGAACTT ATTGATTTAG GTGTTGAAAA CGACATCGTT GaTAAATCAG GAGCATGGTA     780

TTCTTACAAT GGCGAACGAA TGGGTCAAGG TAAGGAAAAT GTTAAAATGT ACTTGAAAGA     840

AAATCCACAA ATTAAAGAAG AAATTGATCG TAAATTGAGA GAAAAATTAG GTATATCTGA     900

TGGTGATGTT GAAGAAACAG AAGATGCACC AAAGTCATTA TTTGACGAAG AATAGTACAC     960

AAATTTATAT CTATAGTTAA ACTTAGCAAA TATCCTTATA GGATTGATTG AAAGTGATAT    1020

TCATCTCATA AAGCTAGAAT AATATCTAAC TTTATGGGAT ACACTACAAA TCGAGACTAT    1080

AAGGTTTTTT ATTTTATTTA TTATTACATT ATCAATAGTT TTATAATCGA GCTTCAAAAC    1140

TTTAGAAAAT AGTAGAAATA GCATTCAATA TAGTGCAAAA GTGCAAATTG ATAACTTGAC    1200

ACTTATCTCC TATAAACCGT ACAATTAATT TGTATGATTT ATATATAATT TCATAAAGTC    1260

ATATTGAATT TCATATAAAG AGCAAACCCT AGAAAAGGAG GTGTTTGTGT GAATTTATTA    1320

AGCCTCCTAC TCATTTTGCT GGGGATCATT CTAGGAGTTG TTGGAGGGTA TGTTGTTGCC    1380

CGAAATTTGT TGCTTCAAAA GCAATCACAA GCTAGACAAA CTGCCGAAGA TATTGTAAAT    1440

CAAGCACATA AAGAAGCTGA CAATATCAAA AAAGAGAAAT TACTTGAGGC AAAAGAAGAA    1500

AACCAAATCC TAAGAGAACA AACTGAAGCA GAACTACGAG AAAGACGTAG CGAACTTCAA    1560

AGACAAGAAA CCCGACTTCT TCAAAAAGAA GAAAACTTAG AGCGCAAATC TGATCTATTA    1620

GATAAAAAAG ATGAGATTTT AGAGCAAAAA GAATCAAAAA TTGAAGAAAA ACAACAACAA    1680
```

```
CGCATCTCCG GTCTCACTCA AGAAGAAGCT ATTAATGAGC AACTTCAAAG AGTAGAGGAA      1800

GAACTGTCAC AAGATATTGC AGTACTTGTT AAAGAAAAAG AAAAAGAAGC TAAAGAAAAA      1860

GTTGATAAAA CAGCAAAAGA ATTATTAGCT ACAGCAGTAC AAAGATTAGC AGCAGATCAC      1920

ACAAGTGAAT CAACGGTATC AGTAGTTAAC TTACCTAATG ATGAGATGAA AGGTCGAATC      1980

ATTGGACGAG AAGGACGAAA CATCCGCACA CTTGAAACTT TAACTGGCAT TGATTTAATT      2040

ATTGATGACA CACCAGAAGC GGTTATATTA TCTGGTTTTG ATCCAATAAG AAGAGAAATT      2100

GCTAGAACAG CACTTGTTAA CTTAGTATCT GATGGACGTA TTCATCCAGG TAGAATTGAA      2160

GATATGGTCG AAAAAGCTAG AAAAGAAGTA GACGATATTA TTAGAGAAGC AGGTGAACAA      2220

GCTACATTTG AAGTGAACGC ACATAATATG CATCCTGACT TAGTAAAAAT TGTAGGGCGT      2280

TTAAACTATC GTACGAGTTA CGGTCAAAAT GTACTTAAAC ATTCAATTGA AGTTGCGCAT      2340

CTTGCTAGTA TGTTAGCTGC TGAGCTAGGC GAAGATGAGA CATTAGCGAA ACGAGCTGGA      2400

CTTTTACATG ATGTTGGTAA AGCAATTGAT CATGAAGTAG AAGGTAGTCA TGTTGAAATC      2460

GGTGTAGAAT TAGCGAAAAA ATATGGTGAA AATGAAACAG TTATTAATGC AATCCATTCT      2520

CATCATGGTG ATGTTGAACC TACATCTATT ATATCTATCC TTGTTGCTGC TGCAGATGCA      2580

TTGTCTGCGG CTCGTCCAGG TGCAAGAAAA GAAACATTAG AGAATTATAT TCGTCGATTA      2640

GAACGTTTAG AAACGTTATC AGAAAGTTAT GATGGTGTAG AAAAAGCATT TGCGATTCAG      2700

GCAGGTAGAG AAATCCGAGT GATTGTATCT CCTGAAGAAA TTGATGATTT AAAATCTTAT      2760

CGATTGGCTA GAGATATTAA AAATCAGATT GAAGATGAAT TACAATATCC TGGTCATATC      2820

AAGGTGACAG TTGTTCGAGA GACTAGAGCA GTAGAATATG CGAAATAATT TTTGTCTCCC      2880

TCACAAATTA GTGAGGGAGC TTTTTTAAGT TGTAGTCTTA AtCTAGTTAG ACAGCACTTT      2940

ATCGGTAATA ACTATATTAA ACAGTAGTTA TTTGAAAGTA AGACGGACCT TATATTAAAT      3000

AAGAAGTTAT TGCTTTTAAT AAAAATGTTT TAGGCTTCGT AATTACTATA TTTATATTAT      3060

GTAAACCTAT AAAGATGATT GGTTTTCTAT CCAATAAAAA AGAAGAGAAG ATGTAACACA      3120

TCTTCTCTTC yGCAATATTA ATTAGGATTT ATTTCTAAGT TGAGTTATTT TAATTGTAAA      3180

TCTGTTTTCT TTAATTCTTT TATAACTTCT GCAGTATCAT AACAATTTGT TGCAATTGTT      3240

GAATATCTCT CTGCTAAACG ATATGCATTA ATGTAAAGCT TTAAACTTTC TTTAGCTATA      3300

TCCTCTGCAT CTTCGAATTT TGATGGGTTA GACATAACCA CTAATTCTGC AAATTTTTCT      3360

GGATCAATAT TAATAGACAT GTATTTATTT ACAACTCCTA TTTATTTTGA TGTCTTAATA      3420

CTAACATATT GAAGTTTTCA GACAAAGTAA TGTCTCTCTA TAATTGAAGA AAAATAATTC      3480
```

```
GGATGAACAA AACATGAGAA TAATGTTTAT AGGGGATATC GTAGGTAAAA TTGGACGAGA      3600

CGCAATTGAA ACGTACATAC CTCAACTGAA GCAAAAGTAT AAACCAACAG TTACAATTGT      3660

AAATGCTGAA AATGCAGCAC ATGGTAAAGG TTTGACTGAA AAAATATATA AACAATTACT      3720

AAGAAATGGT GTAGATTTCA TGACTATGGG TAATCACACA TATGGTCAAC GTGAAATTTA      3780

TGATTTTATA GATGAAGCAA AACGACTAGT AAGACCAGCG AATTTTCCGG ATGAAGCGCC      3840

GGGAATTGGT ATGAGATTTA TACAAATTAA TGATATTAAA CTTGCAGTTA TTAATCTGCA      3900

AGGAAGAGCG TTTATGCCAG ATATTGATGA TCCTTTTAAA AAGGCAGATC AATTAGTCAA      3960

GGAAGCACAA GAACAAACTC CGTTTATATT TGTTGATTTT CATGCAGAAA CAACTTCTGA      4020

AAAGTATGCA ATGGGATGGC ATTTAGATGG TAGAsTAGCG CTGTTGTTGG AACGCATACA      4080

CACATTCAAA CAGCAGATGA ACGTATTTTA CCAAAGGGGA CAGGGTATAT AACGGATGTT      4140

GGTATGACAG GTTTTTATGA TGGCATTTTA GGAATAAATA AAACAGAGGT AATTGAGCGT      4200

TTTATCACTA GTTTGCCACA AAGACATGTT GTTCCAAATG AAGGTAGAAG TGTATTATCT      4260

GGTGTTGTTA TTGATTTAGA CAAAGAAGGT AAAACAAAGC ACATCGAACG TATATTGATA      4320

AATGATGACC ATCCATTTTC AACATTTTAA AATTACGTAA GTAAACATTC GAATTGGACC      4380

CTATCGTCCA TTAGTATGAA TTTAATATAG TACCACTGTT TACATAGTAA ATCGGTGGTT      4440

CTTTTTGTTA TCATTTAATA TGAAATATAT CCATAGGAGG CATATAACTA TGAAACCACA      4500

ATTATCGTGG AAAGTTGGCG GTCAACAAGG CGAAGGTATT GAATCAACTG GGGAAATCTT      4560

CGCTACGGCT ATGAATAGAA AAGGATATTA TTTATATGGA TATAGACATT TTTCAAGTCG      4620

TATCAAAGGT GGACATACGA ATAATAAAAT TAGAGTTTCT ACGACGCCTG TTCATGCAAT      4680

TAGTGATGAT TTAGATATTT TGATTGCATT TGACCAAGAA ACAATTGATG TTAACCATCA      4740

TGAAATGAGA GAAGACAGTA TTATTTTArC TGATGCCAAG GCTAAACCTG TGAAaCCAGA      4800

AGGATGTCAT GCACAGCTTA TTGAATTACC TTTTACAGCA ACCGCTAAAG AATTAGGTAC      4860

AGCATTAATG AAAAACATGG TTGCAATAGG TGCTACTAGC GCATTGATGA ATTTGAATAC      4920

AAATACATTT GAAGAACTTA TTACTAATAT GTTTTCTAAA AAAGGTGACA AGGTAGTTGA      4980

AGTCAATATC CAAGCATTAA ACGAAGGTTA TCAATTAATG CAATCTCGCT TACCTGAAAT      5040

CTACGGGGAC TTTGAATTAG AGTCAACAGA TGCACTACCA CATCTATATA TGATTGGTAA      5100

CGATGCCATT GGATTAGGTG CAATTGCTGC AGGTTCACAA TTTATGGCGG CATATCCTAT      5160

TACACCTGCG TCTGAAGTTA TGGAATATAT GATTGCCAAT ATATCTAAAG TAAACGGAGC      5220

GGTTATTCAA ACAGAAGATG AAAATTGCTGC TGTAACTATG GCTATTGGTG CAAATTATGG      5280
```

```
TGGATTATCT GGTATGACTG AAACGCCATT AGTCATTATT AATACCCAAC GAGGTGGACC   5400

TTCTACTGGA TTACCTACGA AACAAGAACA GTCAGATTTA ATGCAAATGA TTTATGGTAC   5460

ACATGGTGAT ATTCCAAAAA TTGTTGTAGC ACCAACAGAT GCAGAAGATG CATTTTATTT   5520

AACTATGGAA GCATTTAATT TAGCAGAACA ATATCAATGC CCTGTTATAG TTCTAAGTGA   5580

TTTGCAATTA TCTTTAGGTA AACAAACTGT TGAAAAATTA GATTATAATC GTATTGAAAT   5640

TAAACGTGGT GAAATCATTC AATCTGATAT TGAACGTGAA GAAGATGATA AAGGTTATTT   5700

CAAGCGTTAT GCGTtAACAT CCGATGGTGT TTCTCCTAGA CCTATCCCCG GTGTTAAAGG   5760

AGGTATTCAT CATATAACTG GTGTGGAaCa CAATGAAGAA GGTAAACCTA GTGAATCTGC   5820

GTCAAATAGA CAACAACAAA TGGAAAAACG AATGCGTAAA ATTGAGCAGT TACTAATTGA   5880

ATCGCCAGTA GAAGCTAACT TACAACATGA GGATGCAGAT ATTCTTTATA TCGGTTTTAT   5940

TTCTACAAAA GGTGCAATTC AAGAAGGTAG TAACCGTTTG AATCAACAAG GCATAAAAGT   6000

TAACACTATA CAAATTAGAC AATTGCATCC ATTCCCAACA AGCGTTATTC AAGATGCAGT   6060

TAATAAAGCG AAGAAAGTCG TTGTAGTGGA GCACAATTAT CAAGGACAAT TGGCTAGTAT   6120

TATAAAAATG AATGTCAATA TTCATGATAA GATTGAAAAT TATACAAAGT ATGATGGGAC   6180

ACCTTTCCTA CCACATGAAA TCGAAGAAAA AGGCAAAATA ATTGCTACTG AAATAAAGGA   6240

GATGGTATAG ATGGCGACAT TTAAAGATTT TAGAAATAAT GTTAAGCCTA ACTGGTGCCC   6300

CGGATGTGGC GATTTCTCAG TACAAGCTGC AATTCAAAAA GCAGCCGCAA ATATAGGGTT   6360

AGAACCTGAA GAAGTAGCTA TCATCACCGG TATAGGATGT TCTGGCCGTC TTTCAGGATA   6420

TATTAATTCT TATGGCGTTC ATTCTATTCA CGGACGTGCA TTACCTTTAG CTCAAGGTGT   6480

AAAAATGGCG AATAAAGATT TAACTGTTAT TGCATCGGGA GGAGATGGTG ATGGTTATGC   6540

TATAGGTATG GGGCATACAA TCCATGCTTT AAGAAGAAAT ATGAACATGA CGTATATAGT   6600

CATGGATAAT CAAATTTATG GTTTGACAAA GGGACAAACA TCGCCGTCAT CAGCAGTAGG   6660

ATTTGTTACT AAAACAACGC CAAAAGGTAA TATAGAAAAA AATGTTGCGC CTTTAGAATT   6720

AGTATTATCA TCTGGTGCCA CATTTGTAGC CCAAGGTTTT TCAAGCGATA TTAAAGGATT   6780

AACAAAACTA ATTGAAGATG cAATTAATCA TGATGGATTT TCATTCGTTA ATGTCTTTTC   6840

ACCATGTGTG ACTTATAATA AAATTAACAC ATACGATTGG TTTaAAGAAC ATTTAACAAG   6900

TGTTGATGAc ATTGAAAATT ATGATTCTAC AGATAAACAA TTAGCGACTA AAACTGTTAT   6960

TGAACATGAA TCTTTAGTAA CTGGTATTGT TTATCaAGAT AAAGAAACAC CATCATATGA   7020

ATCtCAAATT AAAGAGTTAG ATGATmCACC ACTTGCTAAA AGAGATATCa AAATTaCTGA   7080
```

```
TGTATTTATA ACAGATCCAT TTATGCTACT CAGTTTTTTA CTATTACAAA AAATAAAGGA        7200

GTTTTTAAAA ATGAAAGACA CATTAATGAG TATACAAATA ATTCCTAAAA CACCAAACAA        7260

TGACAATGTT ATACCTTACG TAGACGAGGC GATTAAAATA ATTGACGAAT CTGGTTTGCA        7320

TTTTAGAGTA GGTCCGTTAG AAACGACAGT ACAAGGAAAT ATGAATGAAT GTTTAATTTT        7380

AATACAATCA TTAAATGAAC GAATGGTGGA ACTTGAATGT CCAAGTATTA TTAGCCAAGT        7440

TAAGTTTTAT CATGTGCCAG ATGGCATCAC TATTGAAACT TTAACTGAAA AATATGATGA        7500

ATAACATTAA AAGTGAAGTA AACTGGATTT GAATTGGCTT GTTAGAGATG ACGTATAACT        7560

TTAACTGTTT TTGCACTTTA TAGTTAAATT TAATATAATT ATTAAATGAT ACGGGCAAAT        7620

AGAAAGGATT TTGTAAAGTG AACGAAGAAC AAAGAAAAGC AAGTTCTGTA GATGTTTTAG        7680

CTGAGAGAGA TAAGAAAGCA GAAAAAGATT ATAGTAAATA TTTTGAACAT GTTTATCAGC        7740

CGCCTAATTT AAAAGCAAGC GCAAAAAAAG AGGTnAAA                                7778

(2) INFORMATION FOR SEQ ID NO: 49:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 1128 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear


        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGATGAAGTT GTTACgAAAA TTGCGTACGC TGTTTCAGAA CATGTCAAAA TAGAAACAGG         60

TAATCCATTC TTTCAAACAT CACATAGTGG TTGTGCGACG GGCGGATCCT GTAATTGTTC        120

ATTATAAAAA ACATCGAGTC AGAAAAAGGT GGTTATTGAA cCACTAACTA GCATCTGACT        180

CGATGTTTTT ATTTATTCGG GATTGTTTGT TTGAATTGTT GTGCTAAATC TGGTCGATCT        240

GTCACAATCG TGTGTGCACC TTTTTGGTAT AAATCATTCA TCAGATTTAT ACTATTTACG        300

CCATAATAGC CTGGAATGAT ATTCATATCA TTTAACCATT TGATAAAACG AGATGAAGTC        360

AAATCAATGC CTTTAAAATG AGTAGGCATT TGGAACGTTT GTGCTAATGG TTGGTAGTAC        420

CTACCACCTA ATAAATGATA TTTTAAAAAT GCTTCTGTAA CTTCCTGTTG GCTAGCACCA        480

ATTGCGACGG ATCCTTGTGC AATTTTATTA AAACGAACGA TTTGTTCTTT ATAAAAACTT        540

GTCACAAGAA CGCGGTCAAA TGCTTGATTT TCTGCAATTG TATCAAACAT AATTTGTGGT        600

GCGATTGAGC CTTCATAGGA TTCAGGAGCA TCTTTTAAGT CTACGTTTAT ATACATATCA        660

GGATATTGCT TCAGCAACTc ATCGAAGGTT AGTATAGCTG TGTGTGCATG ACCACGATAT        720
```

391

AATGTATGGG CACTAACTTT TCCAGAGCCG TTCGTCGTTC TATCAACAGT TGCGTCATGA   840

AAAACGATAA GCTGTTGATC TTTTGTGAGT CTCACATCTG TTTCAAAGCC ATCAACGCCT   900

AATTGTTTAG CATAGTCAAA TGCAAGTTGC GTTTGCTCTG GTCTTAAAGC CATACCACCG   960

CGATGCGCAA ATATATATGG TGCATTGCCT TTGAAAAAAG CAGGGATGGT TTGCTTTTTA   1020

GTAATCACTT TATTTTTATT GATCATTAAT AGACTACTTA AAAATCCAGC ACCGACTAGT   1080

ACCGCATTTA AAATGTTTCT GTTTACnTTT TTCATAAAAA ATTCCTCC   .   1128

(2) INFORMATION FOR SEQ ID NO: 50:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6252 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CAAGCAAACA ATCGTCGATA AAATTGCTAA AATAATAAAA GTAATTCGAA CTTTCATCAT   60

GATCATCCTT TGTTTATAGA GTCAATATAA GTATGGAATA TGTTAGGTAT ATAGTCAAAT   120

GCGTCAACTA ATGGGAATTT TGGCATAGAT AGAGAATTTA AGGCAATTAA AAAGGCATCA   180

AACAGTAATA TGCTGCTTGA TGCCCAAATG ATGACTTTAG CTAAATTGAT TAGTCACTTT   240

TAAAGATAAA GAATTGTCAT GAATTAAAAC TCATGTAATG ATGTGTTACA TTTCGCAATG   300

ATGGCTTTCA GTTATTTATC GATAACATCA CTCTTGATAC CTTTAGATTT TAAGAAATCT   360

TTAATTTTAT CTTGTTGCTT TTTATTAACA TCACCGGCAT ATTTTGTTGG CACGTCGACA   420

ACATTGATTT TATTTTGCGG TTGATAGCTA AGCTTTTCAA TATCTTCATC AACATTGGCG   480

ATTGTACTAT TTAAAGCTTT GAAGTAATTC ATCATTAATT CAACGGGTTT CTTATATTCT   540

TTAGGAATAT TGTTTTCAGT GACAAATTTC TTGAAATGCA AATCGTTTTT AACAGCTAAG   600

TTAGATAAGT GGCTAAGTGT TTCTGCTTGT TTTTCAGTCA CTTTTGTTTG ACTGTCAATT   660

TGTTTATCTA GTTTATGTTG CATAATATAT TTGTTATCAA GTATATCGCT ATTTACAGAC   720

AAATACTTTT CTATAGCTTG CTTCATCTCT GCATCACTAA TATCACTATT TTTCTTATCT   780

GAGTTAAAGA TATCTTTTGT tTCTAATTTT TTAGCGCTTT TAGGTGCATG GATGCCAGTA   840

CTTGTATGAT GATCTTCGTT ATCAGATTGA TCGGACGCGC AACCTGTAAG AATTAATGTC   900

GATGCTAAAA ATGTACTTAG TAGTAATCTC TTTTTCATAA TGTAATATAA CTCCTTAGTT   960

TATCTTTAAT TGAAAAAATA TGTATTCATG TTTAATAGAG TAACATTGAA TTAGTTTGGA   1020

```
TCTATCAATA ATGCATCATT TTGGACGTTG TTAAGGATAG CTTTATCTAT AAATAACTGC      1140

ATAATTGGTT GTACTAATTT AGACGTAGGT ATCGTACGTA AAAGCATAAT AATTTCGTTC      1200

ACATACTTTT CTTTCTCAAT ATCATTTTTC ATATTGATTT GTTTGCGAGA GGTACATACT      1260

TTAAGCATTA TCGCACATCT CGTTGTATAT ATTAAGTTTA TCATAACATG ATTTTATGTC      1320

GGGATAAAAA AATAACAGCA TCTTAACAAA TGTAAGATAC TGTCAGTGAA ATGAATGAAA      1380

CTTTAGTTTC TGaTAATATA GTCAAAGGCA TTTAATGCTG CATTTGCACC AGCGCCCATT      1440

GAAATGATAA TTTGTTTGTT CTTCTGATCT GTGACATCGC CAGCAGCAAA TATTCCAGGA      1500

ACATTCGTAT TATTGTTACG ATCAATCACA ATTTCACCAC GTTCGTTTAA TTCAACAGCA      1560

TCGTTTAACC ATGATGTGTT TGGAAGTAAA CCAATTTGAA CAAAGATACC ATCTAAGTTA      1620

AGTAGATGTT CTTCGCCGGT GTTCATGTCT TCGTAACGTA TACCTGTAAC ATGGTCTTCT      1680

CCGACAACTT CAGTAGTTTT GGCATTTGTT TTGATATCAA CATTTGATAA AGAACGTAAA      1740

CGATCTTGTA ACACGTTGTC TGCTTTTAAT TCGCTAGCGA ATTCGAATAA TGTAACATGA      1800

TTAACGATAC CAGCAAGGTC AATTGCTGCT TCAACCCCAG AGTTACCGCC ACCGATAACT      1860

GCTACGTCTT TATTTTCAAA TAGAGGTCCG TCACAGTGAG GGCAGAATGC AACACCTTTA      1920

TTAATCAATT GCTCTTCACC TGGAATGTTT AGCTTACGCC AACCTGCACC AGTAGCAATA      1980

ATGACTGTTT TACTTTCTAA GACAGCACCG TTTTCTAACG TAACTTTAAT TGCTTCGTCA      2040

GTCTTTTCGA TATCTGTAGC ACGTATACCT GTCATTGCAT CAATGTCATA TTGATCAATG      2100

TGCGCTGCTA AGTTAGAAGA AAAATTCAGAA CCAGTTGTTT CTTTAACAGT AATGAAGTTC     2160

TCAATACCAG CAGTATCATT AACTTGGCCA CCGATACGAT CAGCAACTAT ACCAGTACGT      2220

AAACCTTTAC GTGCTGTGTA AATCGCTGCA CTACCACTAG CAGGACCACC ACCAACGATT      2280

AAGaCATCAT AAGGTTCTTT ATTTTCAAAC TCAGATGCAT CTGCCGTACT GCCTAGTTTC      2340

GAAAGAATAT CTTGGATTGT CATACGACCA TTGCCAAATT CTTCGCCATT TAAAAAGACA      2400

GCAGGGACTG CCATGATGTT TTCAGATTCT TCACGGAACA CTGCACCATC AATCATAGAA      2460

TGCGTGATGT TAGGGTTGAT CACACTCATT AAGTTAAGTG CTTGAACGAC ATCAGGACAT      2520

TTTTGACACG TTAAACTAAT GAATGTTTCA AAATGGAATG AACCTTCTAA TTTTTTAATT      2580

TGGTCAATGA TTGACTGTTT TTCTTTAGGT GCACGACCAC TAACCTGTAA AATTGCTAAA      2640

ACAAGTGAGT TAAACTCGTG ACCTAATGGA ATACCTGCAA ATGTTACACC TGTTTCTTCG      2700

CCAGGACGAT TGACTGAGAA ACTTGGTGTA CGTTTTAAAG ATTTTTCAGA AAGAGATAGT      2760

CTAGGTGACA TATCAGTAAT TTCTGTCAAC AAATCTTTAA GTTCTTTGGA TTTATCATCT      2820
```

393

```
TGTTGTTTTA AATCAGCATT AAGCATGGTT GTAATGCCTC CTTAGATTTT ACCTACTAAA    2940

TCTAAACCAG GTTGCAATGT TTTAGCGCCT TCTTCCCATT TAGCTGGGCA TACTTCGCCA    3000

GGGTTTTTAC GAACATATTG AGCTGCTTTG ATTTTGTGAG CTAATGTACT AGCGTCACGG    3060

CCAATTCCGT CAGCGTTAAT TTCAGATGCT TGTACAACAC CGTCTGGGTC GATAATGAAT    3120

GTACCACGTT GAGCTAAACC AGTAGCTTCA TCTAATACAT CAAAATTACG AGTGATTGTT    3180

TGTGATGGGT CACCAATCAT AGTGTAAGTG ATTTTGCTAA TTGCATCTGA ATGGTCATGC    3240

CATGCTTTGT GTACGAAGTG AGTATCAGTT GATACTGAGA ATACATTTAC GCCTAATTTT    3300

TGTAATTCTT CATATTGGTT TTGTAAGTCT TCTAATTCAG TTGGACAAAC GAATGAGAAG    3360

TCAGCAGGAT AGAAGCATAC TACGCTCCAA GAACCTTTTA AATCTTCTTG TGTAACTTCT    3420

TTAAATTGAT CTTTTTTTGG ATCGAAArCT TGCGCTGTAA ATGGTAAGAT TTCTTTGTTA    3480

ATTAATGACA TAAATATCTT CCTCCTAAGA ATTTAAGTAT GAATTAGAAC TATCAATTGA    3540

TTGCGCTTAA TTATAATAAT TCTAATCTCT TAGTTAGCAT TATTACATTT TGATCCAGAA    3600

TAGTCAACTG GATAACTTTG TAAAGTGAAT GATTACTTTT AAAATAAAGA AAGATAATAT    3660

AAAGTGCTTT GATAATGGAT TTTGTAGTTG ATGATTTAAA AGGTTGTGTC TATATTTAAT    3720

ATCTTGATTT TAATGTAAAA AATGTAAAAA AAGAAGATTT GTATTCTCAA CTAAGTCAAC    3780

CTTATTGATA ATGGTATGAG AATATTTGTT CGAGATGGAT GAAGGTAATG AGTGAGAAAC    3840

TGGATTTTTA AAGTATGAGA CAATATTTTA AAAAGTTCAA TTATTAACTT ATAAGCAAAT    3900

AATTGCTATA AAAAAGTTTG GACGTGTACA ATTGCAATAT GAAGATTTTA AATTAATTGT    3960

AAAGTATCGA GGAGTGGGTA ACGTGTCAGA ACATGTATAT AATCTTGTGA AAAAGCATCA    4020

TTCTGTTAGA AAATTTAAGA ATAAACCTTT AAGTGAAGAC GTTGTTAAGA AATTGGTAGA    4080

AGCTGGACAA AGCGCTTCGA CGTCAAGTTT CCTGCAAGCA TACTCAATTA TTGGTATCGA    4140

CGATGAGAAG ATTAAAGAAA ATTTACGAGA AGTTTCTGGA CAACCTTATG TTGTAGAAAA    4200

TGGCTATTTA TTCGTCTTTG TTATTGATTA TTATCGTCAT CATTTAGTTG ATCAACATGC    4260

TGAAACTGAT ATGGAAAATG CATATGGTTC AACGGAAGGT TTGCTAGTAG GTGCAATCGA    4320

TGCAGCATTA GTTGCCGAAA ATATTGCGGT AACTGCTGAA GATATGGGGT ATGGCATTGT    4380

CTTTTTAGGA TCATTAAGAA ATGATGTTGA ACGCGTTCGA GAAATTTTAG ACTTACCTGA    4440

CTATGTCTTC CCGGTATTTG GTATGGCAGT AGGGGAACCc GCAGATGACG AAAATGGTGC    4500

AGCCAAGCCA CGCTTACCAT TTGACCATGT CTTCCATCAT AATAAGTATC ATGCTGATAA    4560

GGAAACACAG TATGCACAAA TGGCAGATTA CGACCAGACA ATCAGCGAGT ACTATGATCA    4620
```

CAAAGCAAGA TTAGATATGT TAGAACAATT GCAAAAATCA GGCTTAATAC AGCGATAgCA   4740

AGATACCAAA ATAACCCGCC CCCCTCTAGC TTAAAATGAT AAGTATAGCT AGAGGGGGCG   4800

GGTATTTCTT GCAATGAATT AGTGTGAAGT TAATGCAGCA TTATCATTTG AATCGAAAGT   4860

ATCTTTATCC CAATGTTTAG TTAACTTGGC GGTACCTGTA CCAGCTAGCA TTGAATCGTT   4920

CACGTTTAAT GCTGTTCTAC CCATGTCAAT CAATGGTTCA ACGGAGATGA GCACGCCGGc   4980

TAAAGCGACT GGCAAGTTTA ACGTTGACAA CACCAATATG GATGCAAATG TAGCCCCGCC   5040

ACCGACGCCA GCAACGCCGA ATGAACTAAT AATCACGACA GCGATTAACG TTACAATAAA   5100

TTGTAAATCA ATTTCTACAT TAGCGACGGG TGCGACCATA ATTGCAAGCA TGGCAGGGTA   5160

AATGCCTGCA CAACCATTTT GTCCAATCGA CAATCCAAAT GTCGCAGCGA AATTGGCAAT   5220

ACCTTCTGGC ACGCCTAGAC GTCTTGTTTG TGTTTGTACA TTCAATGGTA AGGCACCCGC   5280

GCTTGAGCGT GATGTGAATG CAAAGATTAA TACTTCCAAA GTCTTTTTAA CATAGCGAAT   5340

TGGGCTAATA CCTAACAGGC TTAAAATAAT TAAGTGAATG ATATACATCG TAATTAATGC   5400

AGCGTACGAT GCGATTAAGA ATTTTCCTAA AGTCCAAATG GCGCCAAAGT CACTTGTCGA   5460

TAATGTGTTG GCCATAATTG CTAATACACC GTATGGCGTT AAACGTAAGA CGAACGTCAC   5520

AATCGCCATT ACTAGTGAAT AGATAGCGTC AATCGCACGC TTAAGCAATT CACCATGATC   5580

AGGTTGTTTG CGTnTACGCG TAAATAAGCA AATCCTATAA ACGAAGCAAA TATCACGACA   5640

GCAATCGTGG aAGTTGCACG TTGTCCaGTG AAATCTAAGA ATGGATTTTT AGGCAATAAT   5700

TCCAAAATTT GTTGTGGTAA CGTATGTGCT GTTAAATCTT TCGCTTGTTT AGCAATTTCG   5760

CTTCCACGTG CTTGTTCAGC GTTACCAAGG TTAATTGTTG ATGCATCTAA ACCAAACACC   5820

AAGGCATACA CAACACCAAC AATCGCAGCA ATGGTGACAG TGCCAATTAA AAAGATAAAA   5880

ATGAGACTAC CAATTTTAGC AAACTTTTCT CCGATTTGAA TTTTAGTGAA TGCAGCTACA   5940

ATAGAAATGA AAATTAAAGG CATAACAATC ATTTGCAACA ATGCAACGTA ACCTTGTCCG   6000

ACAATGTTGA ACCAGTCACT TGTTGATGTA ATAACATTCG AATGTGTGCC ATAAATAAGA   6060

TGCAATAACA CACCGAATAC TATACCAATC CCTAAAGCTG TAAACACACG TTTCGCAAAA   6120

GATATATGTT TGCGAGCCAT CATGTGCAAT ATTACGATGA AAATCACCAA TACAATAATA   6180

TTAATCAGTG TAAGAAAAGC ATTCATGAAC GTCACTCCTT AAATTTTTGA ATATAATTCC   6240

GACTAGTATG CT   6252

(2) INFORMATION FOR SEQ ID NO: 51:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6730 base pairs

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

```
ATCAAATCnC AAAATATTTA TTAATnAnAA GGGGATTATC CaTGTgAGAA ACAAAGTAAT       60

GCTCTTTTTT TACCTCTTGT GGGTTGAAAA aTGGATCATC AGAGATAGAC TTCTTCTTTT      120

TCGAAGATGA CATTTGATAC TTTAATCTTC TAAAACCATA ACTTGTCGCA TCAAAAATGC      180

CTTCTTGTAC AAGTAAAATC AAAAATATGC TAATAAAAAT AATTAATGAA ACATAAAACA      240

ATATATTTAA ATATGTAATG ATAGTATGGC TATTAAAAAG CCATATAATA AACGTTAATA      300

TTGGCGTTAT TAGTGCCATT CCAAGCCATT TTTTCAACAT TTGATCACTC CCACTTATAG      360

AAAACTCTTA CGCATAGTTT ACATTAAAAT CAGACATTGA GGAATGATTT TTTAATTTCT      420

TCAGCTTTAT TGAAATTCTA AAATCAATCA TTCTTCATTA GTTTAAAGCA AAAAAATATT      480

GATATATAGT AAATATTGTA TATATAATAT TAGTTAAGAT TTCaGAAAAT TTTGAAGGGA      540

ATGGAAATTT AGAAATCGGA ATTTGTTAGA GGAGGGGATT AGATGGGGAA ATATATTTTC      600

AAACGATTTA TTTATATGCT TATTTCTTTA TTTATTATTA TTACAATTAC ATTTTTCTTA      660

ATGAAATTAA TGCCAGGTTC GCCATTTAAC GATGCTAAAT TAAATGCTGA ACAAAAAGAA      720

ATTTTAAATG AAAAATATGG ATTAAATGAT CCTGtAGCTA CGCAgTATTT ACATTATTTA      780

AAAAATGTTG TTACAGGCGA TTTTGGTAAT TCATTCCAGT ATCATAATCA ACCTGTGTGG      840

GATTTGATTA AACCGAGACT ACTACCTTCT TTTGAAATGG GTCTTACAGC AATGTTCaTC      900

GGTGTGATAC TGGGACTTAT TTTAGGTGTT GCAGCAGCTA CTAAACAAAA TTCTTGGGTT      960

GACTATACAA CTACAGTTAT TTCAGTTATT GCAGTATCTG TACCATCTTT TGTACTTGCT     1020

GTACTTTTAC AATATGTATT TGCAGTTAAA TTAAGATGGT TCCCAGTAGC TGGATGGGAA     1080

GGTTTTTCGA CCGCGGTATT ACCGTCACTT GCATTATCTG CAGCTGTTTT AGCAACTGTC     1140

GCCAGATACA TAAGAGCAGA GATGATAGAG GTATTAAGTT CAGACTATAT TTTATTAGCG     1200

AGAGCTAAAG GTAATTCGAC AATGCGTGTA CTTTTTGGAC ATGCACTTAG AAATGCTTTA     1260

ATTCCAATTA TTACAATTAT CGTTCCCATG TTAGCAAGTA TTTTAACAGG CACTTTAACA     1320

ATTGAAAATA TTTTTGGAGT TCCTGGATTA GGGGATCAAT TCGTACGTTC AATTACAACA     1380

AATGATTTCT CAGTAATCAT GGCAATCACA CTATTATTTA GCACACTGTT TATCGTTCT      1440

ATTTTTATTG TAGATATTTT GTACGGTGTG ATAGATCCAC GAATTCGTGT TCCcAAGgAGG    1500

TAAAAAATAA TGGCTGAAAA TAAAAACAAT TTGTCGATTA ACGACGATCA TTCTAATGCA     1560
```

TGAATCAGGA	ACCTGAAATG	CAACGAGAAA	GCAAAAACTT	TTGGCAAGAT	GCTTGGGCTC	1680
AGTTAAAACG	AAATAAGTTA	GCTGTTGTCG	GTATGATAGG	TTTAATTATC	ATTGTAATAT	1740
TTGCTTTTAT	CGGTCCAGTT	ATAAATAAAC	ATGATTATGC	TGAACAAAAT	GTAGAACATA	1800
GAAATCTTCC	GGCAAAAATA	CCTGTATTAG	ACAAAGTTCC	ATTTTTACCT	TTTGATGGTA	1860
AAGATGCAGA	TGGCAAGGAT	GCTTATAAAG	CAGCAAATGC	TAAAGAAAAT	TATTGGTTTG	1920
GTACTGATCA	GTTGGGTCGA	GATTTATGGA	CAAGAACATG	GAAAGGTGCT	CAAATTTCAT	1980
TGTTTATCGG	TGTTGTTGCA	GCGATGTTAG	ATATTTTTAT	TGGTGTTGTA	TATGGTGCGA	2040
TTTCTGGATT	CTTCGGTGGA	CGTGTCGATA	CGATTATGCA	ACGTATACTT	GAAGTCATAG	2100
CATCTATTCC	GAATTTAATT	GTCGTAATTT	TATTTGTATT	AATTTTTGAA	CCATCCATTT	2160
GGACAATTAT	ATTGGCTATG	TCTATCACAG	GCTGGTTAGG	CATGAGCAGA	GTTGTACGTG	2220
GAGAATTTTT	AAAATTAAAA	AATCAAGAGT	TTGTCATGGC	TTCGAAAACA	TTGGGGGCTT	2280
CAAAATTCAA	ATTGATATTT	AAGCATATTT	TACCTAATAC	ATTAGGTGCT	ATCGTGGTTA	2340
CATCAATGTT	TACAGTACCT	AGTGCTATTT	TCTTCGAAGC	ATTTTTAAGT	TTCATTGGTA	2400
TAGGTGTACC	CGCACCTCAA	ACATCGTTAG	GGTCATTAGT	AAATGATGGG	CGCGCAATGT	2460
TATTAATTTA	TCCACATGAA	TTATTTATAC	CAGCAATGAT	TTTAAGTTTA	TTAATTCTAT	2520
TCTTTTACTT	ATTTAGTGAT	GGATTACGTG	ATGCATTTGA	TCCGAAAATG	CGTAAATAAA	2580
AAGGGGGCAT	AGCATATGAC	TGAAAGAATA	TTAGAAGTAA	ATGATTTGCA	TGTTTCCTTT	2640
GATATTACAG	CAGGGGAAGT	GCAGGCAGTG	AGAGGCGTAG	ATTTTTATTT	GAACAAAGGG	2700
GAAACATTGG	CAATTGTTGG	TGAATCAGGT	TCAGGTAAAT	CTGTAACAAC	AAAAGCAATT	2760
ACAAAATTAT	TCCAAGGGGA	CACAGGAAGA	ATTAAAAAGG	GAGAAATTTT	ATTTTTAGGG	2820
GAAGATTTAG	CAAAAAAACC	TGAAAATGAG	TTGATTAAAT	TACGTGGCAA	AGATATTTCA	2880
ATGATCTTTC	AAGATCCAAT	GACATCTTTA	AACCCAACGA	TGCAAATTGG	TAAACAAGTC	2940
ATGGAACCAT	TAATTAAGCA	CAAAAATTAT	AGTAAAGCAC	AAGCTAAAAA	GCGCGCATTG	3000
GAAATACTAA	ATCTTGTAGG	TTTACCAAAT	GCAGAAAAAA	GATTTAAAGC	ATATCCTCAT	3060
CAATTTTCAG	GTGGACAAAG	GCAAAGAATT	GTTATTGCAA	CCGCATTAGC	TTGTGAACCT	3120
AAAGTGCTCA	TTGCTGATGA	ACCAACGACT	GCATTAGACG	TAACGATGCA	GGCACAAATT	3180
TTAGATTTAA	TGAAAGAACT	ACAACAAAAA	ATCGATACAG	CAATTATTTT	TATAACGCAT	3240
GATTTAGGGG	TTGTTGCGAA	TATTGCTGAT	AGAGTGGCAG	TTATGTATGG	TGGTCAAATG	3300
GTTGAAACAG	GAGATGTTAA	CGAAATATTT	TATGATCCAA	AGCATCCATA	TACATGGGGA	3360